Physica A 492 (2018) 1456–1472

Contents lists available at ScienceDirect

Physica A

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

Calibration of short rate term structure models from bid–ask coupon bond prices

Erika Gomes-Gonçalves [a], Henryk Gzyl [c,*], Silvia Mayoral [b]

[a] Independent Consultant
[b] Dept. of Business Administration, Universidad Carlos III de Madrid, Spain
[c] Centro de Finanzas, IESA, Caracas, Bolivarian Republic of Venezuela

HIGHLIGHTS

- We present a method to obtain zero coupon bonds from bid–ask prices.
- The method is non-parametric and model free.
- The method can be extended to solve for the short rate term structure.
- The procedure is based on the method on maximum entropy in the mean.
- We solve compare its performance to traditional methods to value swaps.

ARTICLE INFO

Article history:
Received 15 June 2017
Received in revised form 26 September 2017
Available online 2 December 2017

Keywords:
Zero coupon bond prices
Short rate model calibration
Maximum entropy in the mean
Interest rate swaps

ABSTRACT

In this work we use the method of maximum entropy in the mean to provide a model free, non-parametric methodology that uses only market data to provide the prices of the zero coupon bonds, and then, a term structure of the short rates. The data used consists of the prices of the bid–ask ranges of a few coupon bonds quoted in the market. The prices of the zero coupon bonds obtained in the first stage, are then used as input to solve a recursive set of equations to determine a binomial recombinant model of the short term structure of the interest rates.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction and preliminaries

In order to introduce the necessary notation and to make this note self contained as possible, in the first two sub-sections along with the statement of the problems to solve, we review the basic aspects about interest rates, and in order to describe the examples, we review how the valuation of interest rate derivatives is carried out within the binomial model of interest rates. Then we continue with the statement of the problems to be solved and a description of the contents of the paper.

1.1. First problem: determination of prices from the bid–ask spread

Interest rates, and their term structure are a key input for cash flow valuation and the determination of the price of all sort of financial products. It is therefore important to have ways of relate the term structure of bond prices to the market price of bonds. Before stating the problems to be solved, we shall establish the notation to be used throughout. First of all,

* Corresponding author.
E-mail addresses: erikapat@gmail.com (E. Gomes-Gonçalves), henryk.gzyl@iesa.edu.ve (H. Gzyl), smayoral@emp.uc3m.es (S. Mayoral).

https://doi.org/10.1016/j.physa.2017.11.073
0378-4371/© 2017 Elsevier B.V. All rights reserved.

we shall consider a time horizon of T years which without loss of generality we split into $N = 2T$ semesters, and we suppose that all cash exchanges occur at the end of each semester. Thus, we label time in semester units, so $t = 0, 1, \ldots, N$ denotes the number of half years gone by.

By $P(0, t)$ we shall denote the price at $t = 0$ of a currency unit received at time t. Therefore consistency requires that $P(0, 0) = 1$, and financial intuition requires that $P(0, t + 1) < P(0, t) < 1$. We are excluding the case in which interest rates are negative and that constraint is violated. The price at time 0 of a cash flow (a sequence of payments) $\boldsymbol{F} = \{F(t) : t = 0, 1, \ldots, N\}$ taking place at the end of each half year, also called the present value of the cash flow at $t = 0$ is defined by:

$$PV(\boldsymbol{F}) = \sum_{t=0}^{N} F(t)P(0, t). \tag{1.1}$$

The price at time 0 of a unit of currency to be received at t, $P(0, t)$, is also called "discount factor". The discount factors $P(0, t)$ are determined by the price of borrowing by central banks when they issue coupon bonds. The newcomer to the field may consult the nice book by Luenberger [1] for concepts like cash flow valuation, interest rates and especially, the no arbitrage (fair pricing) argument as to why (1.1) is to be linear in the $f(t)$'s.

Regretfully, central banks do not specify the discount factors $P(0, t)$ for all values of t. But they do it implicitly by issuing bonds with and without coupons and of varying maturities. We denote the maturities of the available bonds by $\{t_1, t_2, \ldots, t_M\}$ measured in semesters, where M is the number of available bonds. The cash flow implicit in their coupons will be denoted by $\boldsymbol{c}^{(i)} = \{c(i, t) : t = 1, \ldots., N\}$ for $i = 1, \ldots, M$. That is, $c(i, t)$ is the coupon paid at the end of the t-th half year by the i-th bond. By convention we suppose that $c(i, t) = 0$ whenever $t > t_i$.

The mathematical problem that this situation generates consists of determining the price of the zero coupon bonds from the information contained in the prices of the bonds and the coupons to be received. The price of the bond (its present value at $t = 0$,) is given by

$$\pi(i) = \sum_{t=1}^{N} c(i, t)P(0, t). \tag{1.2}$$

Since, for example, central banks issue a dozen bonds, and the maturities go from a semester up to 30 years, the problem of determining the $P(0, t)$ from (1.2) happens to be an ill-posed, linear inverse problem, with convex constraints on the solution. The constraints come from the condition $1 = P(0, 0) > P(0, 1) > \cdots > P(0, N) > 0$.

Actually, there may be two added complications to the problem of determining $P(0, t)$ from (1.2). Sometimes the bond prices are listed in the market up to a bid–ask range, and not only that, the prices may be misquoted (or mispriced), and in this case (1.2) has to be replaced by

$$\sum_{t=1}^{N} c(i, t)P(0, t) + \epsilon_i \in (b_i, a_i) \tag{1.3}$$

where (b_i, a_i) is the bid–ask price range. We mention in advance, that sometimes the bid–ask spread may be quite small. Observe that even when the prices are perfectly known ($a_i = \pi_i = b_i$), there might be some inconsistencies among the prices. The ϵ_i's are there to pick these up, and are part of the unknowns in the system (1.3). And the specification of possible ranges $[-d_i, d_i]$ for these errors imposes further constraints upon the solutions to (1.3).

It is one of the aims of this note to show how to determine the $P(0, t)$ and the ϵ_i subject to the mentioned constraints and satisfying Eq. (1.3).

In a previous note, Gzyl and Mayoral [2], we showed how to solve that problem when the prices are known exactly. There we reviewed several methods of solution to the problem based on interpolation by several types of splines, and provided a list of references to these methods. Regretfully we missed some references to other methods of solution. The reader might consider Ronn [3], McKay and Prisman [4] and Ioffe [5] and the references therein to review some of the effort carried out to solve the mathematical problem at hand.

Once the $P(0, t) : t = 1, \ldots, N$ have been determined we shall use them to calibrate a binomial model for the short term structure. But first, for completeness let us recall some facts about the term structure of interest rates and about the binomial model.

1.2. The random interest rates case

The bond price term structure refers to the collection of bond prices specified by $\{P(k, t) \,|\, t = 0, 1, \ldots, T; 0 \leq k \leq t\}$. In the deterministic case, $P(k, t)$ is the price at time k of a unit of currency to be received at time t. These are related to the zero coupon prices by $P(0, t) = P(0, k)P(k, t)$. When $t = k + 1$ the concept of short rate is introduced by setting

$$P(k, k + 1) = \frac{1}{1 + r(k)}.$$

The reason for the notational convention is that the short rate is the interest rate during $[k, k+1]$ is supposed to be known at the beginning of the time interval. The last identity implies

$$P(0,t) = \prod_{k=1}^{t} \frac{1}{1 + r(k-1)}. \tag{1.4}$$

Clearly, if the short term structure is positive, the $P(0, t)$ will be decreasing. All that very nice, useful and extensively used. What is missing in the modeling described above, is the fact that actual short rates, which can be regarded as the building blocks of the term structure, should be considered to be random. Note that at the beginning of any time lapse, the short rate for that semester is supposed to be known, thus the randomness means that the short rates beyond the immediate semester are unknown and are to be described by random variables. So further modeling is called for.

As far as models to describe the randomness goes, the recombinant binomial model is about the simplest model that allows us to capture many features of the intrinsic randomness of the short term rates, and is useful for the computation of prices of interest rate derivatives. Below we shall briefly review the properties of that model. See Luenberger [1] and especially Cairns [6] for more detail and further references. What we eventually aim at, is to use the solution to the problem described in the previous section as input for the determination of the short rates in the framework of the recombinant binomial model.

In this model, the randomness is specified as follows. The underlying randomness is modeled by binary sequences of length N, that is, as underlying probability space we chose $\Omega = \{0, 1\}^N$, and as space of events the class of all subsets of Ω. If ξ_k denotes the result of the k-th toss, the binomial model supposed that the state $s(k)$ of the market at $t = k$ is one of the possible values $\sum_{j=1}^{k} \xi_j$. We shall furthermore suppose that the $\xi_i,\ i = 1, \dots, N$ are independent and identically distributed with $\mathbb{Q}(\xi_i = 1) = 1/2$, which results in $s(k)$ being a $B(k, 1/2)$ random variable.

Let us denote by $r(k-1, j),\ j = 0, \dots, k$ the values of the short rate during the time lapse $[k-1, k]$ if at the beginning of the period the world is in state $(k-1, j)$. At the end of the time interval, one of two things may happen: Either $\xi_k = 1$ and the state at $t = k$ becomes $(k, j+1)$ and the corresponding short rate during the period $[k, k+1]$ becomes $r(k, j+1)$, or $\xi_k = 0$ and the state at $t = k$ becomes (k, j) and the short rate for $[k, k+1]$ is $r(k, j)$. To simplify notation, we propose using $r_k(\mathbf{s}) = r(k, s(k))$ where as said, $s(k) = \sum_{j=1}^{k} \xi_j$ denotes the random state at $t = k$, for $k = 0, \dots, N-1$. Using this notation, the value at time t of one unit of currency deposited at time 0 is one of the possible values of

$$B(0,t) = \prod_{k=1}^{t} \left(1 + r(k-1, s(k-1))\right). \tag{1.5}$$

That is, the value of our savings at a future date will depend on what happens in the market between now and then, and it is to be modeled by a random variable.

In the random case, the term structure $\{P(k, t) \mid t = 0, 1, \dots, T; 0 \le k \le T\}$, is a collection of random processes. To spell out their properties we need to introduce some further notation. Along with the sample space Ω we consider a collection of $\sigma-$algebras $\mathcal{F}_k : k = 0, \dots, N$ defined by $\mathcal{F}_0 = \{\emptyset, \Omega\}$, and $\mathcal{F}_k = \sigma(\xi_1, \dots, \xi_k)$ for $k = 1, \dots, N$. To complete the framework, the standing assumption is that for each $t \le N$, the present value at time 0 of bond prices $P(k, t)$ are a martingale, more explicitly, the following holds

$$\frac{P(k,t)}{B(0,k)} = E_{\mathbb{Q}}\left[\frac{P(k+1,t)}{B(0,k+1)} | \mathcal{F}_k\right]. \tag{1.6}$$

The value at time 0 of a (possibly random) cash flow $\{F(k) : k = 1, \dots, N\}$ is now given by

$$PV(\boldsymbol{F}) = \sum_{k=1}^{N} E_{\mathbb{Q}}\left[\frac{F(k)}{B(0,k)}\right]. \tag{1.7}$$

A variation on the theme of (1.7) is the following. The value (price) at time 0 of a zero coupon bond paying 1 at time t is computed by means of

$$P(0,t) = E_{\mathbb{Q}}\left[\frac{1}{B(0,t)}\right], \tag{1.8}$$

This relates the price of zero coupon bonds at time 0 structure to the short term rates. Note as well that the price of a bet that pays 1 if state j occurs at time t is

$$P_0(t,j) = E_{\mathbb{Q}}\left[\frac{\zeta(t,j)}{B(0,t)}\right], \tag{1.9}$$

where $\zeta(t, j) = I_j(s(t))$, where I_A is the indicator function of the set A. Clearly as $\sum_j \zeta(t, j) = 1$ we have

$$P(0,t) = \sum_{j=0}^{t} P_0(t,j). \tag{1.10}$$

We shall call $P_0(t,j)$ the price at time 0 of the random bet which pays 1 if the market is at state j at time t. The following recurrences play a key role in relating the $P(0,t)$ to the short rates in the recombinant binomial model, see Cairns [6] for further details.

Theorem 1.1. *Suppose that the ξ_j are independent and $\mathbb{Q}(\xi_j = 1) = q = 1 - \mathbb{Q}(\xi_j = 0)$. The following recurrence relations hold:*

$$P_0(t+1,t+1) = qP(t,t+1,t)P_0(t,t),$$

$$P_0(t+1,j) = qP(t,t+1,j-1)P_0(t,j-1) + (1-q)P(t,t+1,j)P_0(t,j), \quad \textit{for} \quad 1 \le j \le t.$$

$$P_0(t+1,0) = (1-q)P(t,t+1,0)P_0(t,0).$$

Here $P(t,t+1,j) = (1+r(t,j))^{-1}$ denotes the discount factor during $[t,t+1]$ if the market is in state j at time t.

Let us now carry out a few calculations to prepare the stage for the recursive calculation of the short term rates. Observe to begin with, that at time $t=0$ the interest rate for the interval $[0,1]$, or equivalently, the price $P(0,1)$ of the bond maturing at the end of that lapse, are known and related by $P(0,1) = (1+r(0,0))^{-1}$.
Suppose now that we have already determined the state prices $P_0(t,j)$. According to (1.1)-(1.8)-(1.9)-(1.10), and the recurrence in Theorem 1.1 we can relate the $P(t,t+1,j)$ and the $P_0(t,j)$ to $P(0,t+1)$ as follows

$$P(0,t+1) = \sum_{j=0}^{t+1} P_0(t+1,j) = (1-q)P_0(t,0)P(t,t+1,0) + qP_0(t,t)P(t,t+1,t)$$
$$+ \sum_{j=1}^{t} \left((1-q)P_0(t,j)P(t,t+1,j) + qP_0(t,j-1)P(t,t+1,j-1)\right).$$

Here, in the first line we wrote the terms with $j=0$ and $j=t+1$ and the remaining ones in the second. This identity can be rearranged into

$$P(0,t+1) = \sum_{j=0}^{t} P_0(t,j)P(t,t+1,j). \tag{1.11}$$

This identity plus (1.8) and (1.9) will be used as the basis for the recursive determination of the $P(t,t+1,j)$. It is interesting to note that, for $t=1$ and due to the fact that $P_0(1,j) = E_Q[I_j(s(1))/B(0,1)] = P(0,1)P(S(1)=j)$, for $t=2$ the identity (1.11) looks as follows,

$$P(0,2) = P(0,1)\left(\frac{1}{2}P(1,2,1) + \frac{1}{2}P(1,2,0)\right).$$

1.3. The problems to solve and description of contents

We have already mentioned that the first goal of this work is to use (1.3) combined with an extension of the standard maximum entropy method to determine the zero coupon bond prices $P(0,t): t = 1,\ldots,N$. With the notations introduced in the previous section, we can state the second more precisely: It consists of using maxentropic techniques to determine the binomial term structure of the short rates, that is determine the $\{r(t,j) : j = 0,\ldots,t;\ t = 0,\ldots,N-1\}$ from the $P(0,t)$ obtained in the first part. Also, and for the sake of comparison, we will also use the $P(0,t)$ to determine the binomial term structure with two parametric models, the Black, Derman and Toy [7] (BDT for short) and, the Ho and Lee [8] (HL for short) models.

Both problems consist of solving linear systems of equations in which there are convex constraints imposed upon the solution, and for which the data is given up to an interval. The method of maximum entropy in the mean (MEM) is an efficient technique to deal with such problems. For the sake of completeness, we present a generic version of this method in Section 2.

In Section 3 we explain how the adapt method to deal with the constraints of problem (1.3). In Section 4, we first recall the parametrization of the BDT and the HL short rate models. Then we explain how the results in Section 1.2 lead to a recursive set of constrained linear problems and how the maxentropic technique can be used to obtain a parameter free determination of the short term rates.

Section 5 is devoted to a description of the numerical results. First we shall examine the case of data given up to an interval and then we shall examine the effect of the noise (or possible mispricing) on the zero coupon price curves.

Then we shall proceed to the second aim of the work, namely, to recover the short rates from the zero coupon data. For the sake of comparison, first we calibrate two standard parametric (the BDT and the HL) models. After that, we carry out the non-parametric calibration using the maxentropic method. Then we compare the outputs of the three procedures. As these three methods provide us with a binomial tree of shot term rates, in order to compare the output of the three procedures, we consider an interest rate swap. We examine how would a swap buyer do if the interest rates were described by the binomial trees obtained by the parametric and the non-parametric methods. We close with a few concluding remarks.

2. The method of maximum entropy in the mean

In this section we establish the basic formalism of the MEM. To begin, let us consider (1.7) in a context free notation. The problem that we want to solve is to find $\boldsymbol{x} \in \mathcal{C} \subset \mathbb{R}^N$ and $\boldsymbol{y} \in \mathcal{D} \subset \mathbb{R}^M$ such that

$$\boldsymbol{Dx} + \boldsymbol{y} \in \mathcal{K}_d$$

where $\boldsymbol{D}$ is an $M \times N-$ matrix, and the sets $\mathcal{C}$, $\mathcal{D}$ and $\mathcal{K}_d$ are supposed to be convex in their ambient spaces and to have non-empty interiors. As a matter of fact, we shall eventually consider $\mathcal{C} = [0, 1]^N$, $\mathcal{D} = [-d, d]^M$ for some positive d, and $\mathcal{K}_d = \prod_i^M [b_i, a_i]$. To further simplify, let us introduce the augmented objects

$$\boldsymbol{A} = [\boldsymbol{DI}], \quad \boldsymbol{z} = \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix} \quad \mathcal{K}_s = \mathcal{C} \times \mathcal{D}$$

and restate our problem as consisting of finding $\boldsymbol{z} \in \mathcal{K}_s$ such that

$$A\boldsymbol{z} \in \mathcal{K}_d. \tag{2.1}$$

Think of the subscripts "s" and "d" to stand for "signal" and respectively "data".

The idea behind MEM is rather simple: Consider an auxiliary random variable $\boldsymbol{Z}$, taking values in $\mathcal{K}_s$ such that $\boldsymbol{z} = E_P[\boldsymbol{Z}]$ with respect to an unknown probability measure P that is supported by $\mathcal{K}_s$. If such probability measure can be found, the constraint upon $\boldsymbol{z}$ is automatically satisfied. To define $\boldsymbol{Z}$ we consider the probability space $(\mathcal{K}_s, \mathcal{B}(\mathcal{K}_s), Q)$, where $\mathcal{B}(\mathcal{K}_s)$ denotes the Borel subsets of $\mathcal{K}_s$, and Q is any ($\sigma-$ finite) measure such that the convex hull $con(supp(Q))$ generated by its support equals $\mathcal{K}_s$. What this last requirement achieves for us, is that any probability P having a positive density $\rho(\boldsymbol{z})$ with respect to Q, will satisfy $E_P[\boldsymbol{Z}] \in \mathcal{K}_s$. The standard maximum entropy (SME) procedure enters as a stepping stone at this stage. On the class of densities (with respect to Q) we define the entropy function by

$$S(\rho) = -\int_{\mathcal{K}_s} \rho(\boldsymbol{z}) \ln \rho(\boldsymbol{z}) dQ(\boldsymbol{z}),$$

or $-\infty$ if the integral of $|\ln \rho(\boldsymbol{z})|$ is not convergent. The choice of sign is conventional, we want to maximize entropy. The entropy function is strictly concave due to the strict concavity of the logarithm. To find a $\rho^*(\boldsymbol{z})$ such that $dP^* = \rho^* dQ$ satisfies (2.1) we solve the problem

$$\text{Find} \quad \rho^* \quad \text{at which} \quad \sup\{S(\rho) | \boldsymbol{A}E_P[\boldsymbol{Z}] \in \mathcal{K}_d\} \quad \text{is achieved} \tag{2.2}$$

The maximization can be split into a sequence of two steps:

$$\sup_{\boldsymbol{m} \in \mathcal{K}_d} \sup\{S(\rho_{\boldsymbol{m}}) | \boldsymbol{A}E_P[\boldsymbol{Z}] = \boldsymbol{m}\}$$

in which the inner maximization has a standard solution given by

$$\rho_{\boldsymbol{m}}(\boldsymbol{z}) = \frac{e^{-\langle \boldsymbol{\lambda}^*, \boldsymbol{Az} \rangle}}{Z(\boldsymbol{\lambda}^*)} \tag{2.3}$$

where the normalization factor is clearly given by

$$Z(\boldsymbol{\lambda}) = \int_{\mathcal{K}_s} e^{-\langle \boldsymbol{\lambda}^*, \boldsymbol{Az} \rangle} dQ(\boldsymbol{z}).$$

The $\boldsymbol{\lambda}^*$ appearing in (2.3) is to be determined minimizing the convex function (dual entropy) given by

$$\Sigma(\boldsymbol{\lambda}, \boldsymbol{m}) = \ln Z(\boldsymbol{\lambda}) + \langle \boldsymbol{\lambda}, \boldsymbol{m} \rangle$$

over $\{\boldsymbol{\lambda} \in \mathbb{R}^M \mid Z(\boldsymbol{\lambda}) < \infty\}$, which in all cases of interest for us will be $\mathbb{R}^M$ itself. Furthermore, and very important,

$$S(\rho^*_{\boldsymbol{m}}) = \Sigma(\boldsymbol{\lambda}^*, \boldsymbol{m}).$$

Therefore, the double maximization described above becomes

$$\sup_{\boldsymbol{m} \in \mathcal{K}_d} S(\rho^*_{\boldsymbol{m}}) = \sup_{\boldsymbol{m} \in \mathcal{K}_d} \inf\{\Sigma(\lambda, \boldsymbol{m}) \mid \lambda \in \mathbb{R}^M\} = \inf_{\lambda \in \mathbb{R}^M} \{\ln Z(\boldsymbol{\lambda}) + \sup_{\boldsymbol{m} \in \mathcal{K}_d} \{\langle \boldsymbol{\lambda}, \boldsymbol{m} \rangle\}\}. \tag{2.4}$$

To make the pending computations explicit, note that

$$\sup_{\boldsymbol{m}\in\mathcal{K}_d}\{\langle\boldsymbol{\lambda},\boldsymbol{m}\rangle\} = \sum_{i=1}^{M}\left(\frac{a_i-b_i}{2}|\lambda_i| + \frac{a_i+b_i}{2}\lambda_i\right).$$

And we still must specify how to compute $Z(\boldsymbol{\lambda})$. For that, consider the following measure Q on $\mathcal{K}_s = [0,1]^N \times [-d,d]^M$:

$$Q(d\boldsymbol{z}) = \prod_{j=1}^{N}\left(\epsilon_0(dx_j)+\epsilon_1(dx_j)\right)\prod_{i=1}^{M}\left(\epsilon_{-d}(dy_i)+\epsilon_d(dy_i)\right).$$

We use the notation $\epsilon_a(dx)$ to denote the measure that puts a unit mass at the point a. It takes a simple computation to verify that

$$Z(\boldsymbol{\lambda}) = \prod_{j=1}^{N}\left(1+e^{-(\boldsymbol{D}^t\boldsymbol{\lambda})_j}\right)\prod_{i=1}^{M}\left(e^{d\lambda_i}+e^{-d\lambda_i}\right),$$

where the superscript "t" denotes transposition and $(\boldsymbol{D}^t\boldsymbol{\lambda})_j = \sum_{i=1}^{M}\lambda_i D_{i,j}$.
With all this, the dual entropy to be minimized in (2.4) becomes

$$\Sigma(\boldsymbol{\lambda},\boldsymbol{m}) = \ln Z(\boldsymbol{\lambda}) + \sum_{i=1}^{M}\left(\frac{a_i-b_i}{2}|\lambda_i| + \frac{a_i+b_i}{2}\lambda_i\right) \tag{2.5}$$

Once the $\boldsymbol{\lambda}^*$ that minimizes the right hand side of (2.5) has been found, then

$$\begin{cases} x_j^* = \dfrac{e^{-(\boldsymbol{D}^t\boldsymbol{\lambda}^*)_j}}{1+e^{-(\boldsymbol{D}^t\boldsymbol{\lambda}^*)_j}} & j=1,\ldots,N \\[2ex] y_i^* = \dfrac{de^{-d\lambda_i^*}-de^{d\lambda_i^*}}{e^{d\lambda_i^*}+e^{-d\lambda_i^*}} & i=1,\ldots,M \end{cases} \tag{2.6}$$

That this is so becomes clear as we consider the derivatives of $\ln Z(\boldsymbol{\lambda})$ with respect to the λ_i evaluated at $\boldsymbol{\lambda}^*$.

3. Application of MEM to determine the zero coupon prices

We make use of a slight change of notation to restate (1.7) into the form of (2.1). That is due to the constraint $P(0,0) = 1 > P(0,1) > \cdots > P(0,N)$. We add a first row and a first column to the coupon price matrix $\boldsymbol{C}$ – but we denote the augmented matrix by the same name – as follows. The new $(M+1)\times(N+1)-$matrix has new first row and columns defined by $C(0,0)=1$, $C(0,j)=0$ for $j=1,\ldots,N$, and $C(i,0)=0$ for $i=1,\ldots,M$. Thus, M will still stand for the number of non-trivial bonds and N for the non-trivial bond prices to be determined. Correspondingly, the vector of coupon prices has a first component equal to 1, and the remaining N equal to $P(0,t)$. Let $\boldsymbol{T}$ be the square upper triangular $(N+1)-$matrix with components $T(t,j)=1$ for $t\le j$ and 0 otherwise. Let $P(0,t)=\sum_{j=0}^{N}T(t,j)x_j = \sum_{j=t}^{N}x_j$, for $t=0,\ldots,N$. Clearly, now we have $\boldsymbol{x}\in[0,1]^{(N+1)}$. To finish, let the matrix $\boldsymbol{D}$ be defined by $\boldsymbol{D}=\boldsymbol{CT}$.

To complete the translation table, we must specify the constraint sets. We already mentioned that $\mathcal{C}=[0,1]^{N+1}$. Since the extra constraint ($P(0,0)=1$) has no possible mispricing and no bid–ask price range, we set $\mathcal{D}=0\times[-d,d]^M$, therefore $\mathcal{K}_s=\mathcal{C}\times\mathcal{D}$, and to finish, $\mathcal{K}_d=\{1\}\times\prod_{i=1}^{M}[b_i,a_i]$. Note as well that since the first constraint on the vector $\boldsymbol{x}$ is $\sum x_j = 1$, the analogue of the dual entropy to be minimized is

$$\Sigma(\boldsymbol{\lambda},\boldsymbol{m}) = \ln Z(\boldsymbol{\lambda}) + \lambda_0 + \sum_{i=1}^{M}\left(\frac{a_i-b_i}{2}|\lambda_i| + \frac{a_i+b_i}{2}\lambda_i\right) \tag{3.1}$$

where now

$$Z(\boldsymbol{\lambda}) = \prod_{j=0}^{N}\left(1+e^{-(\boldsymbol{D}^t\boldsymbol{\lambda})_j}\right)\prod_{i=1}^{M}\left(e^{d\lambda_i}+e^{-d\lambda_i}\right),$$

This completes the description of the framework for the application of MEM to solving (1.7).

Taking all these comments into account and taking care of the relabeling of the range if the subscripts the corresponding version of (2.6) is now

$$\begin{cases} x_j^* = \dfrac{e^{-(\boldsymbol{D}^t\boldsymbol{\lambda}^*)_j}}{1+e^{-(\boldsymbol{D}^t\boldsymbol{\lambda}^*)_j}} & j=0,\ldots,N \\[2ex] y_i^* = \dfrac{de^{-d\lambda_i^*}-de^{d\lambda_i^*}}{e^{d\lambda_i^*}+e^{-d\lambda_i^*}} & i=1,\ldots,M \end{cases} \tag{3.2}$$

4. Calibration of the short rate term structure

Here we rapidly recall the parametric representation of two common models for binomial short rates, after which we explain how to determine the $P(t, t+1, j)$ using (1.11) as starting point. For that we use the monotonicity constraint implied by the two BDT and HL models as starting point, and then use a bound on the variance to determine a range for the unknowns.

4.1. Two standard models for the short rates

There are two simple and useful parametric models that are used within the recombinant binomial model framework: The Black–Derman–Toy and the Ho–Lee models. They prescribe the following parametric representations for the short rates:

$$\begin{aligned} \text{BDT model} \quad & r(t,j) = a_t e^{jb_t}, \quad t = 1,\ldots,N; \quad j = 0,\ldots,t. \\ \text{HL model} \quad & r(t,j) = a_t + jb_t, \quad t = 1,\ldots,N; \quad j = 0,\ldots,t. \end{aligned} \tag{4.1}$$

This leads to the representations

$$\text{BDT model} \quad P(t,t+1,j) = \frac{1}{1+a_t e^{jb_t}}, \quad t = 1,\ldots,N; \quad j = 0,\ldots,t.$$

$$\text{HL model} \quad P(t,t+1,j) = \frac{1}{1+a_t + jb_t}, \quad t = 1,\ldots,N; \quad j = 0,\ldots,t.$$

Note that, as $r(t, 0) > 0$, we require $a_t > 0$. In both cases the parameter b_k is related to the volatility and, in the simplest approach to calibration, it is supposed to be constant, and then the problem reduces to determining the a_t's. Here we shall apply a genetic algorithm to determine the coefficients using (1.11). Also, as the notation implies, the short rates are supposed to be the semiannual rates. In the tables reported in section of numerical results, we rescale them to be annual rates.

4.2. Maxentropic calibration of the short rates

The starting point is to regard (1.11) as a collection of equations (one for each t) to be solved recursively for the $P(t, t+1, j)$ whenever the $P_0(t, j)$ are known. As these consist of one (linear) equation with $t + 1$ unknowns, constrained as explained in the next paragraph, the method of maximum entropy in the mean is technique to be tried.
Note as well that in the two standard models described above the interest rate increases as j increases, therefore to stay within that convention, we impose the constraint $1 > P(t, t + 1, 0) > \cdots > P(t, t + 1, t) > 0$ on the $P(t, t + 1, j)$. To simplify notation, let us introduce $y_k = P(t, t + 1, t + 1 - k)$ for $k = 1, \ldots, t + 1$, With this we rewrite (1.11) as

$$P(0,t+1) = \sum_{k=1}^{t+1} P_0(t,k-1)P(t,t+1,k-1)$$

$$= \sum_{k=1}^{t+1} P_0(t,t+1-k)P(t,t+1,t+1-k) \equiv \sum_{k=1}^{t+1} P_0(t,t+1-k)y_k$$

Observe now that the inequalities go in the same ordering as the labels, that is, $y_1 < \cdots < y_{t+1}$. Now we will provide bounds on the y_k and take care of the constraints before solving that equation using the standard maximum entropy procedure.

Defining the symbol $D(t)$ by means of the no arbitrage identity $P(0, t + 1) = P(0, t)D(t)$, after dividing the last identity by $P(0, t)$ we can rewrite it as

$$D(t) = \sum_{k=1} q_k y_k \qquad \text{where} \qquad q_k = \frac{P_0(t,t+1-k)}{P(0,t)}, \tag{4.2}$$

where the $\sum_{k=1}^{t+1} q_k = 1$ on account of (1.10). Thus, we can think of the q_k as probabilities. In order to control the range (and the variability of the y_k's), we will choose $\gamma > 1$ and set $U = D(t)^{1/\gamma}$ and $L = D(t)^{\gamma}$, and impose the constraint

$$L < P(t,t+1,j) < U \qquad \text{for all} \qquad j = 0,\ldots,t, \tag{4.3}$$

then, on account of Popoviciu's inequality, the variance of the $P(t, t+1, j)$ or that of the y_k's, be it with respect to the binomial distribution or with respect to the q_k's, is bounded above by $(U - L)^2/4$. Thus the choice of the γ is related to the estimation of the volatility of the short rates at each time.

To take care of the constraint $L < y_1 < \cdots < y_{t+1} < U$ we introduce variables x_j, $j = 1, \ldots, t+2$ such that $0 < x_j < L-U$ for $j = 1, \ldots, t + 2$ and such that

$$y_k = \sum_{j=1}^{k} x_j + L, \qquad k = 1,\ldots,t+1, \qquad \text{and} \qquad \sum_{j=1}^{t+2} x_j = U - L.$$

Table 1
US government bond data.

Maturity (Years)	Coupon	Price
2	0.375	99.8907
3	0.5	99.7599
5	1.375	99.4768
7	1.875	99.6286
10	1.75	99.4534
30	2.875	99.9272

It is in the last identity where the role of the slack variable x_{t+2} becomes apparent. Besides all of that, we set $q_{t+2} = 0$, and rewrite (4.2) and the last identity as the system

$$D(t) - L = \sum_{k=1}^{t+2} q_k \sum_{j=1}^{k} x_j = \sum_{j=1}^{t+2} M_j x_j$$
$$U - L = \sum_{j=1}^{t+2} x_j.$$

Above we set $M_j = \sum_{k=j}^{t+2} q_k$. Now we can get rid of the last equation by setting $z_j = x_j/(U-L)$ and the problem to be solved is: Find $0 < z$ such that $\sum_j z_j = 1$, and

$$\sum_{j=1}^{t+2} M_j z_j = \frac{D(t) - L}{U - L}. \tag{4.4}$$

Since z_j are positive and their sum add up to 1, this problem can be solved using the standard method of maximum entropy. From the results in Section 3, the solution to problem (4.4) is

$$z_j^* = \frac{e^{-\lambda^* M_j}}{Z(\lambda^*)}. \tag{4.5}$$

Now $Z(\lambda) = \sum_{j=1}^{t+2} e^{-\lambda M_j}$, and λ^* is determined minimizing the dual entropy

$$\Sigma(\lambda) = \ln Z(\lambda) + \lambda \frac{D(t) - L}{U - L}.$$

Once the z_j^* are at hand, we have that for $j = 0, \ldots, t$,

$$x_j^* = (U - L) z_j^* \quad \Rightarrow y_k^* = \sum_{j=1} X_j^* + L \quad \text{and} \qquad P^*(t, t+1, j) = y_{t+1-j}^*$$

5. Numerical examples

In this section we carry out the numerical implementation of the various techniques proposed above. We consider the one of example used in Gzyl and Mayoral [2], on one hand, to have it as reference, on the other, because the underlying interest rates at that date were not that low. The data is presented in Table 1 right below.

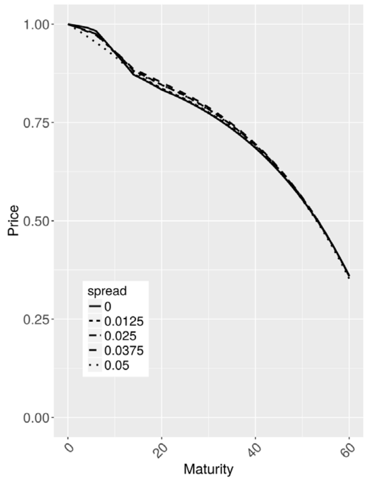
The data was obtained from Thompson-Reuters' web page on July 2, 2013. The coupons are paid semiannually.

5.1. Zero coupon bond prices from possibly noisy bid–ask data

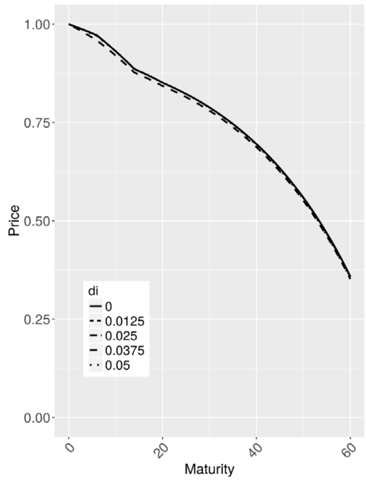
We present three groups of results. One in which there are several possible bid–ask spreads about the quoted prices, other example in which the price is known up to a bid-spread range and possible mispricing is suspected. In the last example, we examine the effect of the bid–ask spread for a fixed mispricing consists in supposing that there may be some mispricing.

In Fig. 1 we present the three sets bond price curves. The curves in Panel (a) were obtained supposing no mispricing ($d_i = 0$, or no error in the data), but a varying bid–ask spread, calculated as $(1 \pm s/2)\pi_i$, where the spread s was set to $\{0, 0.0125, 0.025, 0.035, 0.05\}$. The zero spread curve is plotted in continuous line. We considered a spread as large as 5% to examine the way this affects the zero coupon prices. The conclusion from this numerical example is that the effect of the spread is small. This is actually related to the robustness of the maximum entropy method.

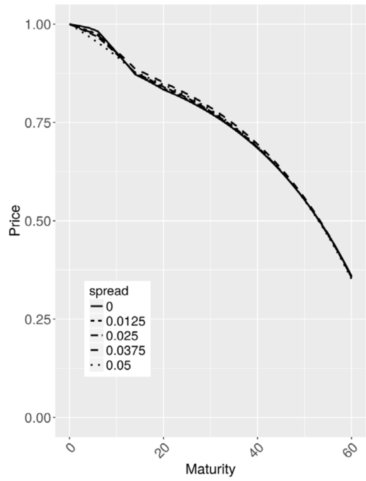
Panel (b) corresponds to the case in which there is a fixed bid–ask (0.0375) spread, and we tested for the influence of several possible ranges for the mispricing. In this case, the supposition of mispricing (or of noise added to the prices) makes the curves to become closer.



(a) Varying bid–ask, no error.



(b) Fixed bid–ask, varying error.



(c) Varying bid–ask, fixed error.

Fig. 1. Zero coupon bond prices from bid–ask data with and without error.

Table 2
RMSE.

Method	RMSE
Ho–Lee	3.2574×10^{-6}
BDT	5.6114×10^{7}
Maxent ($\gamma = 1.5$)	3.53×10^{-8}
Maxent ($\gamma = 2.5$)	1.11×10^{-6}
Maxent ($\gamma = 3.5$)	5.68×10^{-7}
Maxent ($\gamma = 5.0$)	5.03×10^{-7}

Table 3
No arbitrage swap rates.

Ho–Lee	0.0073501
Black–Derman–Toy	0.0073508
Maxent ($\gamma = 1.5$)	0.00735066
Maxent ($\gamma = 2.5$)	0.00735073
Maxent ($\gamma = 3.5$)	0.00735072
Maxent ($\gamma = 5$)	0.00735067

Table 4
Swap distribution.

Model	Prop. positive	Prop. negatives	mean	deviation
Ho–Lee	0.4786	0.5214	−0.00050126	0.0098285
BDT	0.4119	0.5881	−0.00069291	0.0032272
MAXENT ($\gamma = 1.5$)	0.439	0.561	−0.00054	0.00752
MAXENT ($\gamma = 2.5$)	0.419	0.581	−0.00076	0.01790
MAXENT ($\gamma = 3.5$)	0.387	0.613	−0.00091	0.02586
MAXENT ($\gamma = 5.0$)	0.367	0.633	−0.00108	0.03570

For the curves displayed in Panel (c) we allow for the existence of mispricing, but kept the range of mispricing fixed and set to $d = 0.025$. Then we examined the effect of varying the bid–ask spread. Notice now when there is a possible mispricing, the price has similar dependence on the spread, which can perhaps be put as that the presence of noise does reduce the price indeterminacy.

5.2. Calibration of short rates

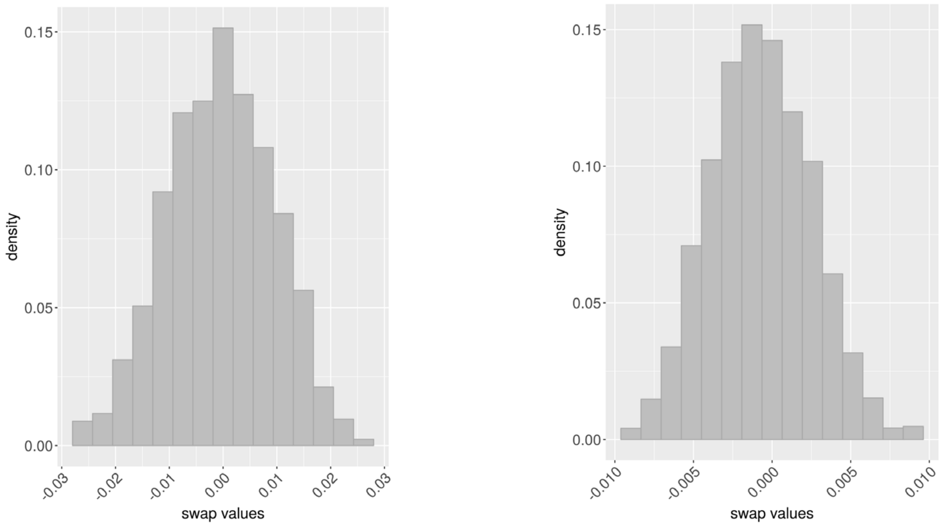
We used the zero coupon bond prices corresponding to the zero bid–ask spread and no mispricing as input for the calibration of the binomial trees according to the BDT and the HL models, as well as the binomial trees using a maxentropic

Table 5
Values of a_t and b_t for the BDT model.

t	0	1	2	3	4	5	6	7	8	9	10	11
b_t	0.181991	0.165914	0.069245	0.035523	0.050383	0.098840	0.017560	0.002278	0.0564388	0.046502	0.022944	0.113037
a_t	0.003114	0.002919	0.002792	0.003086	0.005156	0.005048	0.008701	0.011526	0.012642	0.010271	0.010283	0.011939

Table 6
State prices term structure for the BDT model.

11												0.0004413
10											0.0008970	0.0048678
9										0.0018246	0.0089915	0.0244057
8									0.0036963	0.0164501	0.0405542	0.0734094
7								0.0074889	0.0296267	0.0659094	0.1083822	0.1471904
6							0.0152136	0.0525250	0.1038824	0.1540332	0.1900691	0.2065690
5						0.0306248	0.0913569	0.1578658	0.2081261	0.2314027	0.2285452	0.2070542
4					0.0616137	0.1532269	0.2285750	0.2635674	0.2605896	0.2317423	0.1908252	0.1482305
3				0.1236956	0.2466106	0.3066536	0.3050040	0.2639998	0.2088022	0.1547129	0.1092470	0.0742769
2			0.2483539	0.3713056	0.3701409	0.3068480	0.2289261	0.1586451	0.1045593	0.0663955	0.0410413	0.0248110
1		0.4984476	0.4968521	0.3715130	0.2469049	0.1535185	0.0916378	0.0529590	0.0299173	0.0166205	0.0091361	0.0049722
0	1	0.4984476	0.2484982	0.1239031	0.0617609	0.0307221	0.0152839	0.0075760	0.0037448	0.0018490	0.0009151	0.0004529
$P(0, t)$	1.0000000	0.9968952	0.9937043	0.9904174	0.9870310	0.9815939	0.9759972	0.9646270	0.9529448	0.9409404	0.9286038	0.9166813
t	0	1	2	3	4	5	6	7	8	9	10	11



(a) Ho–Lee.

(b) Black–Derman–Toy.

Fig. 2. Values of the swap for the parametric models binomial short rate tree.

Table 7
$P(t, t+1, k)$ for the BDT model.

11												0.96025
10											0.98723	0.96434
9										0.98463	0.98752	0.96803
8									0.98053	0.98532	0.98780	0.97135
7								0.98842	0.98158	0.98598	0.98807	0.97433
6							0.99042	0.98845	0.98257	0.98660	0.98834	0.97701
5						0.99179	0.99059	0.98848	0.98351	0.98721	0.98860	0.97942
4					0.99373	0.99256	0.99075	0.98850	0.98440	0.98778	0.98885	0.98158
3				0.99658	0.99404	0.99326	0.99091	0.98853	0.98525	0.98833	0.98910	0.98352
2			0.99680	0.99670	0.99433	0.99389	0.99107	0.98855	0.98604	0.98885	0.98935	0.98525
1		0.99657	0.99702	0.99681	0.99461	0.99446	0.99122	0.98858	0.98680	0.98935	0.98959	0.98681
0	0.99690	0.99709	0.99722	0.99692	0.99487	0.99498	0.99137	0.98861	0.98752	0.98983	0.98982	0.98820
t	0	1	2	3	4	5	6	7	8	9	10	11

Table 8
Values of a_t and b_t for the HL model.

t	0	1	2	3	4	5	6	7	8	9	10	11
b_t	0.00067	0.00155	0.00021	0.00024	0.00162	0	0.00283	0.00270	0.00015	0.00113	0.00158	0.00053
a_t	0.00311	0.00288	0.00177	0.003119	0.005068	0.001698	0.011783	0.002374	0.001972	0.012626	0.007396	0.004904

calibration. As we carried out the reconstruction for the 30 year period, the binomial tables are too big. At the end we report only part of the tables and refer the reader to[1] for the full set of tables for the 60 semester time lapse.

The calibration of the two parametric models was carried as follows. Note from the recursive set that once we have determined the $P_0(t, j)$ we can determine the $P(t, t+1, j)$ and then use the recursive set to determine the $P_0(t+1, j)$. For the HL and BDT models specified in the set (4.1), we used a genetic algorithm to determine the parameters that minimize the quadratic errors computed as follows

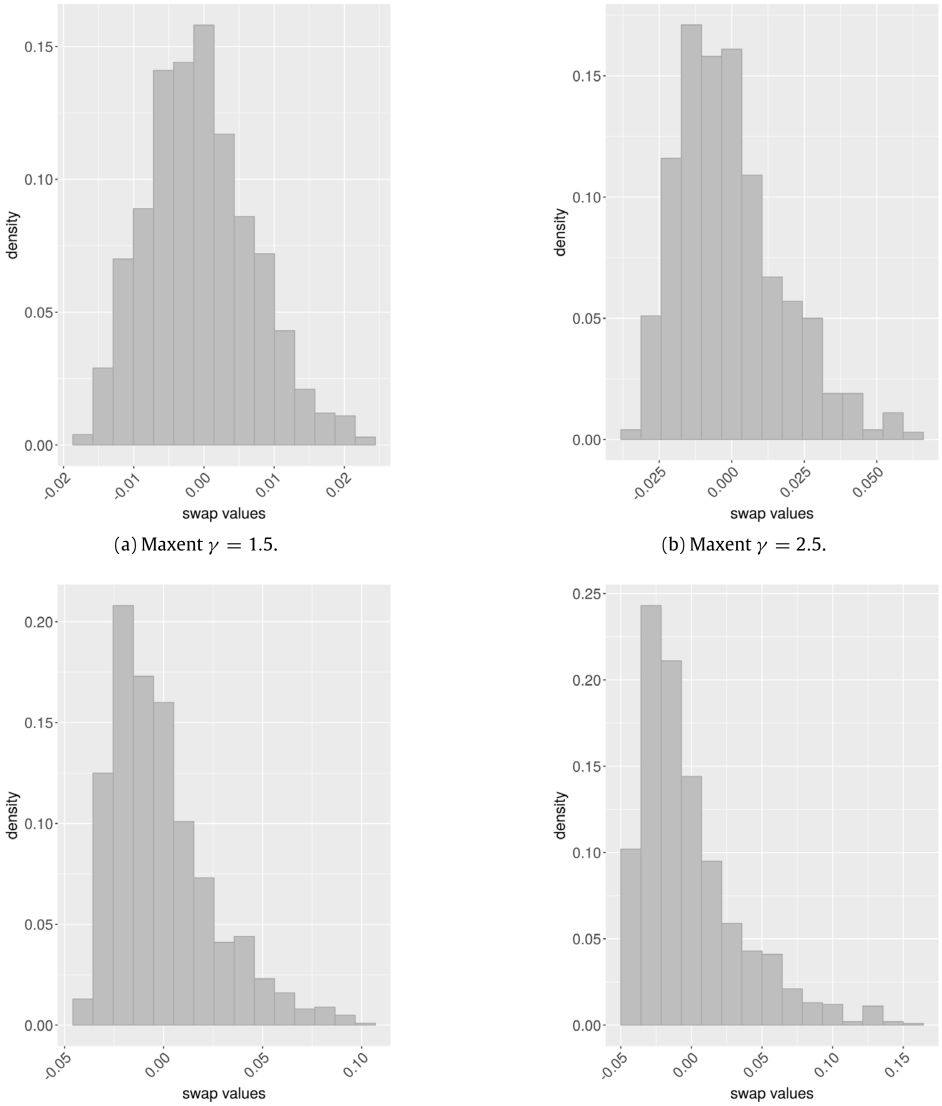
$$\sqrt{\frac{1}{60}\sum_{t=1}^{60}(P(0,t)-P_{inf}(0,t))^2}$$

where P_{inf} is obtained from the inferred state prices $P_0(t, j)$ according to (1.10), that is $P_{inf}(0, t) = \sum_{j=0}^{t} P_0(t, j)$ for each of the cases. The errors are listed in Table 2 and are a measure of the quality of the process. Clearly, the errors are quite small.

1 https://github.com/erikapat/MAXENT_CALIBRATION_RESULTS.

Table 9
State price term structure for the HL model.

11												0.0004329
10											0.0008820	0.0047910
9										0.0017886	0.0088690	0.0241000
8									0.0036617	0.0161945	0.0401322	0.0727384
7								0.0074860	0.0294139	0.0651681	0.1076139	0.1463602
6							0.0151485	0.0525021	0.1033718	0.1529760	0.1893722	0.2061498
5						0.0305935	0.0910929	0.1578066	0.2075946	0.2308503	0.2285129	0.2074043
4					0.0615548	0.1531299	0.2282391	0.2635128	0.2605631	0.2322473	0.1914899	0.1490485
3				0.1235702	0.2464882	0.3065851	0.3049965	0.2640159	0.2093112	0.1557701	0.1100341	0.0749788
2			0.2483435	0.3711743	0.3701365	0.3069108	0.2292572	0.1587121	0.1050884	0.0671639	0.0414937	0.0251455
1		0.4984476	0.4968521	0.3716385	0.2470275	0.1536184	0.0919074	0.0530052	0.0301496	0.0168931	0.0092725	0.0050599
0	1	0.4984476	0.2485085	0.1240345	0.0618244	0.0307563	0.0153521	0.0075866	0.0037843	0.0018884	0.0009324	0.0004628
$P(0, t)$	1.0000000	0.9968952	0.9937041	0.9904175	0.9870314	0.9815940	0.9759936	0.9646273	0.9529387	0.9409404	0.9286048	0.9162392
t	0	1	2	3	4	5	6	7	8	9	10	11



(a) Maxent $\gamma = 1.5$.

(b) Maxent $\gamma = 2.5$.

(c) Maxent $\gamma = 3.5$.

(d) Maxent $\gamma = 5$.

Fig. 3. Values of the swap for the maxentropic models of binomial short rate tree.

Table 10
$P(t, t+1, k)$ for the Ho–Lee model.

11												0.98937
10											0.97736	0.98989
9										0.97773	0.97887	0.99041
8									0.99686	0.97881	0.98038	0.99093
7								0.97914	0.99701	0.97989	0.98190	0.99145
6							0.97203	0.98174	0.99716	0.98097	0.98342	0.99198
5						0.99830	0.97471	0.98435	0.99730	0.98206	0.98495	0.99250
4					0.98859	0.99830	0.97741	0.98698	0.99745	0.98315	0.98648	0.99302
3				0.99619	0.99017	0.99830	0.98012	0.98962	0.99759	0.98424	0.98802	0.99355
2			0.99782	0.99642	0.99176	0.99830	0.98285	0.99228	0.99774	0.98534	0.98956	0.99407
1		0.99559	0.99802	0.99666	0.99336	0.99830	0.98559	0.99495	0.99789	0.98643	0.99111	0.99459
0	0.99690	0.99713	0.99823	0.99689	0.99496	0.99830	0.98835	0.99763	0.99803	0.98753	0.99266	0.99512
t	0	1	2	3	4	5	6	7	8	9	10	11

Table 11
The $P_0(t, j)$ tree by maxentropic calibration $\gamma = 1.5$.

11												0.0004574
10											0.0009233	0.0050136
9										0.0018642	0.0092034	0.0249795
8									0.0037629	0.0167310	0.0412811	0.0746632
7								0.0075933	0.0300318	0.0667311	0.1097088	0.1487448
6							0.0153191	0.0530525	0.1048509	0.1552295	0.1912890	0.2073647
5						0.0307700	0.0917919	0.1588374	0.2091412	0.2320673	0.2286280	0.2064074
4					0.0618004	0.1536795	0.2291485	0.2641440	0.2606526	0.2312097	0.1896828	0.1466896
3				0.1239327	0.2469978	0.3069828	0.3050408	0.2634902	0.2078333	0.1535101	0.1078678	0.0729454
2			0.2485334	0.3715505	0.3701559	0.3065605	0.2283693	0.1576567	0.1035380	0.0654981	0.0402417	0.0241752
1		0.4984476	0.4968523	0.3712761	0.2465178	0.1530442	0.0911659	0.0523932	0.0294664	0.0162976	0.0088943	0.0048062
0	1	0.4984476	0.2483189	0.1236582	0.0615594	0.0305569	0.0151617	0.0074608	0.0036682	0.0018021	0.0008845	0.0004343
$P_{(0,t)}$	1.000000	0.996895	0.993705	0.990418	0.987031	0.981594	0.975997	0.964628	0.952945	0.940941	0.928605	0.916681
t	0	1	2	3	4	5	6	7	8	9	10	11

Table 12
$P(t, t+1, k)$ by maxentropic calibration with $\gamma = 1.5$.

11												0.98937
10											0.97736	0.98989
9										0.97773	0.97887	0.99041
8									0.99686	0.97881	0.98038	0.99093
7								0.97914	0.99701	0.97989	0.98190	0.99145
6							0.97203	0.98174	0.99716	0.98097	0.98342	0.99198
5						0.99830	0.97471	0.98435	0.99730	0.98206	0.98495	0.99250
4					0.98859	0.99830	0.97741	0.98698	0.99745	0.98315	0.98648	0.99302
3				0.99619	0.99017	0.99830	0.98012	0.98962	0.99759	0.98424	0.98802	0.99355
2			0.99782	0.99642	0.99176	0.99830	0.98285	0.99228	0.99774	0.98534	0.98956	0.99407
1		0.99559	0.99802	0.99666	0.99336	0.99830	0.98559	0.99495	0.99789	0.98643	0.99111	0.99459
0	0.99690	0.99713	0.99823	0.99689	0.99496	0.99830	0.98835	0.99763	0.99803	0.98753	0.99266	0.99512
t	0	1	2	3	4	5	6	7	8	9	10	11

In Tables 5 and 8 shown at the end, we list respectively, the values of the parameters (a_t, b_t) for the first 11 semesters. In Tables 6 and 7 presented there, we report as well part of the binomial tree structure for the BDT model, while in Tables 9 and 10 we report the corresponding binomial trees for the HL model.

For the computation of the binomial trees using maximum entropy we used four values of the parameter γ introduced in Section 4.2, namely $\gamma \in \{1.5, 2.5, 3.5, 5\}$. Part of the binomial trees for the case $\gamma = 1.5$ are shown in Tables 11 and 12 at the end.

In the first of the three sets of two tables we list the binomial tree of state prices $P_0(t, j)$, and at the bottom of each column we list the sum of those values, which must coincide with the $P(0, t)$.

5.3. Application to interest rate swaps

Once that the binomial tree of short rates is available, it can be used to answer many questions about interest rates derivatives. Here we consider swaps. These are interesting financial derivatives, that allow their holder to exchange a random cash flow for a constant (or a non-random) cash flow. The idea is that for the investor paying the constant cash flow, there will be no risk involved, or all uncertainty about the future payments has been transferred to the swap seller. For example a corporation paying a debt may want to enter in an interest rate swap, to change a sequence of payments at the actual interest rate for a sequence of payments at a fixed rate.

There are two questions of interest, the answer to which will be obtained from the two stages of the work carried out above. The first is: What is the fair interest rate? In the simplest case, the answer to this question is obtained with the term structure of the bond prices obtained from the market prices of the bonds. The second question concerns the risks for the swap buyer: What is the probability of loss? Or in general, how are the values of the swap distributed. This question can be answered with the aid of the binomial tree of short rates.

To illustrate both issues, we consider the simplest of all cases in which the contract is arranged at time $t = 0$ and the cash exchanges begin at the end of the first semester. Let S be the total number of semesters that the swaps lasts. The fair value of the interest rate is to be chosen so that the expected present value of net cash flow, computed as

$$E[\sum_{t=1}^{S} \frac{1}{B(0,t)} (r_t(\boldsymbol{s}) - r)]$$

equals zero. Since $1 + r_t(\boldsymbol{s}) = (P(t, t+1))^{-1} = B(0, t)B(0, t-1)^{-1}$, with the aid of (1.8) and the fact that $P(0, 0) = 1$, we rewrite the fairness condition as

$$1 - P(0, S) = r \sum_{t=1}^{S} P(0, t) \quad \Rightarrow \quad r = \frac{1 - P(0, S)}{\sum_{t=1}^{S} P(0, t)}.$$

For numerical purposes we shall consider $S = 10$, that is a 5 year interest rate swap, starting $t = 0$, and the swaps occurring at the end of each semester. Also, to make use of the binomial trees that we created to compute the no arbitrage swap rates, we adopted the following point of view: Suppose that all that we know is the binomial rate tree and the binomial state price tree. From that we compute the zero coupon price according to (1.10), and then the no arbitrage rate as indicated above. The resulting swap rates are listed in Table 3.

But more important than computing the fair prices is to foresee what may happen if we enter in such a swap. It is here where the difference in the trees is revealed. For the numerics we need the binomial short rate tree, drawn up to 10 semesters. With the aid of the binomial tree it is easy to check whether $P(t, t+1, j)$ is bigger or smaller than $1/(1 + r)$ during a given semester.

The value of the swap at the end of a possible trajectory (or market history) $\boldsymbol{s}$ of S steps is given by

$$\sum_{t=1}^{S} \frac{1}{B(0,t)} \left(r_t(\boldsymbol{s}) - r \right).$$

in the notation introduced above in Section 1.2. To obtain the distribution of these values, we generate a large number (10,000) of random trajectories of $S = 10$ steps, and compute this last quantity along each trajectory and plot their histogram.

We carried out these computations for the binomial trees determined using the BDT and the HL models, as well as the maxentropic trees for $\gamma \in \{1.5, 2.5, 3.5, 5\}$. The results are summarized Table 4, in which we list the simple statistics of the values as well as the proportion of losses implied by each model. Regarding the proportion of losses, note that the BDT model is more consistent with the maxentropic model for the smaller values of γ whereas the HL model is more consistent with the maxentropic model for larger values of γ. The intuition behind the comparison is that when the swap has less arbitrage then the proportion of times that the buyer wins is equal to fraction of times that the buyer losses. The standard deviations listed only describe which histograms are more concentrated about their means.

In Fig. 2 we present the histograms of the swap values for the two parametric models.

And in Fig. 3 we present the histogram of the distribution of swap values according to the binomial rate tree obtained by maxentropic calibration.

6. Concluding remarks

Clearly the procedures based on the maximum entropy techniques provide us with an interesting model free, and therefore non parametric method, to determine zero coupon bond prices as well as the binomial tree of short rates. The information that is used consists of the market bid ask price of the bonds, and the methodology allows for corrections due to possible mispricing.

That the maxentropic methodology could be used to obtain quite good modeling of the short rates came in as a nice surprise to us as well.

Acknowledgments

Our presentation improved considerably after incorporating all the comments made by the reviewer. We thank her/him for the detailed reading of the manuscript.

References

[1] D. Luenberger, Investment Science, Oxford U. Press, Oxford, 1998.
[2] H. Gzyl, S. Mayoral, Determination of zero coupon and spot rates from treasury data by maximum entropy, Physica A 456 (2016) 38–50.
[3] E.I. Ronn, A new linear programming approach to bond portfolio management, J. Fin. Quant. Anal. 22 (1987) 439–466.
[4] E. Prisman, A unified approach to term structure estimation: A methodology for estimating the term structure in a market with frictions, J. Fin. Quant. Anal. 25 (1990) 127–142.
[5] I. Ioffe, Arbitrage bounds in markets with noisy prices and the puzzle of negative option prices implicit in bonds. J. Bank. Fin. 26, 1199–1228.
[6] A.J.G. Cairns, Interest Rate Models: An Introduction, Princeton U. Press, Princeton, 2004.
[7] F. Black, E. Derman, W. Toy, A one-factor model of interest rates and its application to treasury bond options, Financ. Anal. J. 46 (1990) 33–39.
[8] S.Y. Ho, S.B. Lee, Term structure movements and pricing interest rates contingent claims, J. Finance 41 (1986) 1011–1029.