Applied Mathematical Finance **8**, 27–48 (2001)

Calibrating the Black–Derman–Toy model: some theoretical results

PHELIM P. BOYLE, KEN SENG TAN and WEIDONG TIAN

University of Waterloo, Ontario, Canada, N2L 3G1

Received June 2000

The Black–Derman–Toy (BDT) model is a popular one-factor interest rate model that is widely used by practitioners. One of its advantages is that the model can be calibrated to both the current market term structure of interest rate and the current term structure of volatilities. The input term structure of volatility can be either the short term volatility or the yield volatility. Sandmann and Sondermann derived conditions for the calibration to be feasible when the conditional short rate volatility is used. In this paper conditions are investigated under which calibration to the yield volatility is feasible. Mathematical conditions for this to happen are derived. The restrictions in this case are more complicated than when the short rate volatilities are used since the calibration at each time step now involves the solution of two non-linear equations. The theoretical results are illustrated by showing numerically that in certain situations the calibration based on the yield volatility breaks down for apparently plausible inputs. In implementing the calibration from period n to period $n + 1$, the corresponding yield volatility has to lie within certain bounds. Under certain circumstances these bounds become very tight. For yield volatilities that violate these bounds, the computed short rates for the period $(n, n + 1)$ either become negative or else explode and this feature corresponds to the economic intuition behind the breakdown.

Keywords: interest rate models, Black–Derman–Toy model, volatility, short term, yield

1. Introduction

The modern approach to the modelling of stochastic interest rates started with the classic paper by Vasicek (1977). The early models postulated a stochastic differential equation for the evolution of the short rate of interest and, by invoking no-arbitrage arguments, developed expressions for the prices of pure discount bond and other securities of interest such as options. Other examples of such models include Brennan and Schwartz (1979) and Cox, Ingersoll and Ross (1985). One of the problems with these models was that they did not have enough degrees of freedom to match the model prices of pure discount bonds with the corresponding market prices. This was an uncomfortable situation

The authors are grateful to Baoyan Ding, Fred Guan, Houben Huang, George Lai, David Li, Chonghui Liu, and Ken Vetzal for useful discussion. Phelim P. Boyle thanks the Social Science and Humanities Research Council of Canada and the Natural Sciences and Engineering Research Council of Canada, for research support. Ken Seng Tan acknowledges the research support from the Natural Sciences and Engineering Research Council of Canada. Weidong Tian thanks CITO (Communications and Information Technology Ontario) for research support.

Applied Mathematical Finance ISSN 1350-486X print/ISSN 1466-4313 online
http://www.tandf.co.uk/journals
DOI: 10.1080/13504860110062049

since the model could not be described as arbitrage free when the model prices differed from the current market prices.

One way to resolve this problem is to make the parameters of the short term rate time dependent. In the Vasicek case, the resulting model is known as the extended Vasicek model. This approach has been popularized by Hull and White (1993). Another approach is to model the uncertainty by assuming a stochastic process for the evolution of the forward rate. Since this approach starts from the current observed forward rate, the market prices of today's zero-coupon bonds are built directly into the foundations of the model. The first paper to apply this approach was Ho and Lee (1986) in a discrete-time framework. Subsequently Heath, Jarrow and Morton (1990) (HJM) provided a much more extensive and rigorous approach in continuous-time framework. The HJM model is quite general; it can be calibrated to current bond prices and option prices. In practice, the model parameters are selected by fitting model prices to the current market prices of the most liquid instruments. From a trading perspective, this approach is useful since the model reproduces market prices over for the most liquid traded securities.[1]

For many pricing applications, it is convenient to have a simple binomial type model that fits the current term structure of bond prices and the current term structure of volatilities. In this connection, the Black, Derman and Toy (1990) (BDT) model is a widely used model that can be calibrated to match the term structure of zero-coupon bond prices and the term structure of volatilities. In the original BDT paper, the authors used the yield volatilities as the input term structure of volatility. The yield volatility corresponds to the volatility of the yields on long term bonds. In practical applications, it is often more convenient to use the term structure of short rate volatilities since they can be directly inferred from the market prices of interest rate caps.

Thus, there are two ways to calibrate the BDT model. The first one is the short rate volatility method which uses

- the current term structure of zero coupon bond prices;
- the term structure of future short rate volatilities.

The second one is the yield rate volatility method which uses

- the current term structure of zero coupon bond prices;
- the term structure of yields on zero coupon bonds.

For each of these approaches it is of interest to investigate the conditions under which fitting the BDT model results in a *reasonable calibration*. Our conditions for a reasonable calibration are quite weak. We will require that all the short rates and all the output volatilities[2] in the calibrated BDT model are positive. Sandmann and Sondermann (1993) have analysed the case when the BDT calibration is based on the short rate volatilities. They provide necessary and sufficient conditions for this to happen. Their result is: if the current implied forward rates are all positive (i.e. the pure discount bond price declines as the time to maturity increases) and the short rate volatilities are all positive, then it is possible to calibrate the BDT model. Their conditions are simple from a mathematical perspective and have an intuitive economic interpretation.

[1] This approach has disadvantages from an econometric perspective since, by fitting new parameters on a daily basis, we are effectively assuming a new model every day.

[2] The output volatility under the first approach is the yield volatility as computed from the calibrated BDT model since the input volatility is the short rate volatility. The output volatility under the second approach is the short rate volatility as computed from the calibrated BDT model since the input volatility is the yield rate volatility.

This paper investigates the conditions under which we can calibrate the BDT model when the yield volatility is used. We derive the precise mathematical conditions which the input data must satisfy so that the BDT model can be calibrated. These conditions are less elegant than in the case when the short rate volatility is used. However, we find that it is not possible to calibrate the BDT model for seemingly plausible input term structures. To obtain our conditions we use results from the theory of polynomial equations.

The outline of the rest of the paper is as follows. In the next section, we review the details of the procedure that is used to calibrate a BDT model. We examine both the original and modified BDT models which calibrate to the term structure of yield volatilities and short rate volatilities respectively. We show that the calibration equations can be reduced to a system of polynomial equations so that we can draw on results from the so-called Quantifier Elimination; an algebraic approach that provides conditions for polynomials to have real roots. In Section 3, we provide a detailed analysis of the calibration of a three-period BDT model. We provide both necessary and sufficient conditions for the calibration to be feasible. The conditions are quite complicated even in the three period case and it appears difficult to extend this type of analysis to the n periods case. Hence in Section 4, we provide a sufficient condition for calibration to be feasible from the nth step to the $(n + 1)$th step given that the calibration was successful for the preceding n periods. In Section 5 we provide several examples which illustrate the conditions developed in Section 4. The final section concludes the paper.

2. Calibration of the BDT model

In this section, we review the procedure used to calibrate a BDT interest rate model. In this model, a recombining binomial lattice is constructed so that it matches the current yield curve and the current yield volatility curve. We assume the calibrated binomial lattice has N periods and each period is of size Δt years. Hence the total time horizon of the binomial lattice is $T = N\,\Delta t$ years. The recombining nature of the binomial lattice ensures that at time period n, there are $n + 1$ states. We label these states as $i = 0, 1, \ldots, n$. Let $r(n, i)$ be the (annualized) one-period short rate at period n and state i. The short rate $r(n, i)$ evolves either to $r(n + 1, i)$ (i.e. down-state) or to $r(n + 1, i + 1)$ (i.e. up-state) one period hence with equal risk-neutral probability.

Let $\hat{Y}(0, n) \equiv \hat{Y}(n)$ be the current (market) yield on a n-period zero-coupon bond (i.e. with maturity $n\Delta t$) and $\hat{\sigma}_Y(n)$ be the corresponding current yield volatility. Then the price of an n-period zero-coupon bond, $\hat{P}(0, n) \equiv \hat{P}(n)$, is given as

$$\hat{P}(n) = [1 + \hat{Y}(n)\Delta t]^{-n} \tag{1}$$

Similarly, let $P(n)$ and $Y(n)$ denote the model price and model yield of an n-period zero-coupon bond and $\sigma_Y(n)$ denote the volatility corresponding to the n-period yield implied from the model. In other words, the set $\{P(n), Y(n), \sigma_Y(n)\}$ is similar to $\{\hat{P}(n), \hat{Y}(n), \hat{\sigma}_Y(n)\}$ except that the first set of values is computed from the model while the second set is the market inputs. We calibrate the model to the market by ensuring that $P(n) = \hat{P}(n)$ and $\sigma_Y(n) = \hat{\sigma}_Y(n)$ for all $n = 1, 2, \ldots, N$.

Let us denote the two possible yield realizations at period 1 (i.e. nodes (1, 0) and (1, 1)) on a zero-coupon bond which matures at the end of period n by $Y_d(n)$ and $Y_u(n)$. In the BDT model, these two yields are related by

$$Y_u(n) = Y_d(n) \exp\left[2\sigma_Y(n)\sqrt{\Delta t}\right] \tag{2}$$

In a similar manner, let $P_d(n)$ and $P_u(n)$ denote the prices of zero-coupon bonds corresponding to the yields $Y_d(n)$ and $Y_u(n)$, respectively. Therefore, $P_d(n)$ and $P_u(n)$ represent bonds with $(n-1)\Delta t$ years to maturity and are related to $Y_d(n)$ and $Y_u(n)$ as follows

$$P_d(n) = \frac{1}{[1 + Y_d(n)\Delta t]^{n-1}}$$

$$P_u(n) = \frac{1}{[1 + Y_u(n)\Delta t]^{n-1}}$$

We also have the following relationship:

$$P(n) = \frac{1}{2[1 + r(0,0)\Delta t]}[P_d(n) + P_u(n)] \tag{3}$$

As pointed out by Jamshidian (1991), the calibration procedure is facilitated by the use of a forward induction technique. This involves using the Arrow–Debreu securities, which are defined as follows. Assume we have a security which pays one unit at time n in state i and zero elsewhere and let $A(n, i)$ denote the price at node (0, 0) of this Arrow–Debreu security. The Arrow–Debreu security is sometimes referred to as the Green's function because of its continuous-time analogue.

The Arrow–Debreu prices satisfy the following recursive relation:

$$A(n,i) = \begin{cases} \dfrac{A(n-1,i-1)}{2[1 + r(n-1,i-1)\Delta t]}, & i = n \\ \dfrac{A(n-1,i-1)}{2[1 + r(n-1,i-1)\Delta t]} + \dfrac{A(n-1,i)}{2[1 + r(n-1,i)\Delta t]}, & i = 1, 2, \ldots, n-1 \\ \dfrac{A(n-1,i)}{2[1 + r(n-1,i)\Delta t]}, & i = 0 \end{cases}$$

The model price of an n-period zero-coupon bond can be written in terms of Arrow–Debreu prices as

$$P(n) = \sum_{i=0}^{n} A(n,i) \tag{4}$$

without having to work backwards through the lattice one period at a time to obtain the required value.[3]

Let $A_d(n, i)$ denote the Arrow–Debreu price at node (1, 0) of a contingent claim that pays \$1 if state i is realized in period n and zero otherwise. Similarly, let $A_u(n, i)$ denote the corresponding price of the Arrow–Debreu security at node (1, 1). Then, $P_d(n)$ and $P_u(n)$ can be computed from $A_d(n, i)$ and $A_u(n, i)$ as

$$P_u(n) = \sum_{i=0}^{n} A_u(n,i)$$

[3] More generally, if $X(n, i)$ denotes the payoff of a European contingent claim at node (n, i), then the price of the contingent claim at node (0, 0) is conveniently computed as $\Sigma_{i=0}^{n} A(n, i)\, X(n, i)$.

$$P_d(n) = \sum_{i=0}^{n} A_d(n,i)$$

Note that $A_d(n, n)$ and $A_u(n, 0)$ are zero for all n.

In calibrating a $(N + 1)$-period binomial lattice, the task reduces to finding the values of $r(n, i)$, for $n = 0, 1, \ldots, N$, $i = 0, 1, \ldots, n$, for which the model values are consistent with the input market values. Normally this is carried out one time step at a time. For instance, in the $(n + 1)$th period calibration, the task is to find $r(n, i)$, $i = 0, 1, \ldots, n$ so that the resulting lattice matches the input $\hat{Y}(n + 1)$ and $\hat{\sigma}_Y(n + 1)$, assuming that all the earlier short rates, $r(m, i)$, $m = 0, 1, \ldots, n$, $i = 0, \ldots, m$ have already been calibrated to the input term structures $\{\hat{Y}(m), \hat{\sigma}_Y(m), m = 0, 1, \ldots, n\}$. In other words, the $(n + 1)$th period calibration involves solving the following non-linear equations:

$$\alpha_{n+1} = \hat{P}(n + 1) = P(n + 1) = \sum_{i=0}^{n} \frac{A(n,i)}{1 + r(n,0)(\beta_n)^i \Delta t} \tag{5}$$

$$u^{-1} - 1 = (v^{-1} - 1)\eta_{n+1} \tag{6}$$

where

$$u^n = \sum_{i=0}^{n} \frac{A_u(n,i)}{1 + r(n,0)(\beta_n)^i \Delta t} \tag{7}$$

$$v^n = \sum_{i=0}^{n} \frac{A_d(n,i)}{1 + r(n,0)(\beta_n)^i \Delta t} \tag{8}$$

$$\eta_{n+1} = e^{2\hat{\sigma}_Y(n+1)\sqrt{\Delta t}} \tag{9}$$

$$u^n + v^n = 2\frac{\alpha_{n+1}}{\alpha_1} \tag{10}$$

Note that the non-linear Equations 5 and 6 contain only two unknowns, $r(n, 0)$ and β_n. These equations have to be solved numerically for $r(n, 0)$ and β_n and methods such as the Newton–Raphson iteration approach are often used. Once these unknowns are determined, the short rates in other states are computed via the following recursive relationship among the short rates in each period:

$$r(n,i) = r(n,0)(\beta_n)^i$$

for $i = 1, 2, \ldots, n$.

For $n = 1$ or 2, the calibration is simple. The required short rates are determined as

$$r(0,0) = \hat{Y}(1) \tag{11}$$

$$r(1,0) = \frac{(1 - \eta_2)(1 - 2\frac{\alpha_2}{\alpha_1})\Delta t + \sqrt{D}}{4\frac{\alpha_2}{\alpha_1}\eta_2(\Delta t)^2} \tag{12}$$

$$r(1,1) = r(1,0)\eta_2 \tag{13}$$

where

$$D = \left[(1 - \eta_2)\left(1 - 2\frac{\alpha_2}{\alpha_1}\right)\Delta t\right]^2 + 16\frac{\alpha_2}{\alpha_1}\eta_2(\Delta t)^2\left(1 - 2\frac{\alpha_2}{\alpha_1}\right)$$

For these short rates to be positive and not exceeding the maximum rate ω, the following conditions on the market input must be satisfied:

$$\hat{Y}(1) \leq \omega \tag{14}$$

$$\frac{\omega\Delta t(1 + \eta_2) + 2\eta_2}{2\omega\Delta t(1 + \omega\Delta t + \eta_2) + 2\eta_2} \leq \frac{\alpha_2}{\alpha_1} \leq 1 \tag{15}$$

To verify that most reasonable term structures satisfy the second inequality, we first note that the left-hand most expression is an increasing function of η_2 and converges to $(2 + \omega\Delta t)/2(1 + \omega\Delta t)$ as $\eta_2 \to \infty$. Suppose f_1 is the one-period forward rate from period 1 to period 2; i.e. $\alpha_2(1 + f_1\Delta t) = \alpha_1$. Then any f_1 that satisfies the the following inequality

$$0 \leq f_1 \leq \frac{\omega}{2 + \omega\Delta t}$$

also satisfies (15). In practical situations, setting $\omega = 1$ (or 100%) would be a very conservative upper bound. Also Δt is usually less than 1 year. If we assume $\omega = 1$ and $\Delta t = 1$, the upper bound on f_1 is 1/3. This implies that as long as the forward rate f_1 is positive and does not exceed 33%, inequality (15) is satisfied and the calibration is feasible for any positive input $\hat{\sigma}_Y(2)$. This provides a justification that most reasonable input term structures satisfy inequalities (14) and (15).

It is nontrivial to extend the calibration to one more period; i.e. from period 2 to period 3.[4] In the next section we will analyse the calibration issues for the three-period BDT model.

We now explain a general method to handle the non-linear Equations 5 and 6 using elimination theory. The original system of non-linear equations over the two unknown variables $r(n, 0)$ and β_n can be expressed as a system of non-linear equations over the four unknowns $r(n, 0)$, β_n, u and v. More precisely, Equations 5, 6, 7 and 8 can be expressed as four polynomial equations. This allows us to invoke an important set of results from classical algebra known as Quantifier Elimination (QE). The QE provides a methodology to solve polynomial equations. The underlying principle can be summarized as follows: suppose we are given several polynomial equations. To see if a given polynomial equation has real solutions, it suffices to check if the coefficients of the polynomial equations satisfy certain conditions. A more precise statement of this is given in Appendix A. As a familiar illustration, let us consider the quadratic equation $ax^2 + bx + c = 0$. In this case, we only need to check whether the coefficients (a, b, c) satisfy the relation $b^2 - 4ac \geq 0$ for real solutions to exist.

We now consider a simple example using QE. Suppose we are interested in finding criteria for the existence of real solutions $r(n, 0)$ in the range (a_1, a_2), where $a_1 < a_2$. Let us introduce two new variables x and y, so that $r(n, 0) = a_1 + x^2$ and $r(n, 0) = a_2 - y^2$. The elimination theory provides conditions among the coefficients of the polynomial equations such that $r_{n,\,0} \in (a_1, a_2)$ if and only if the coefficients satisfy certain relations.

Although elimination theory is constructive in the sense that there are algorithms for finding the relations among the coefficients, all algorithms are either impractical or very complicated to implement.

[4] Our colleague, Ken Vetzal pointed out that extensions of results from $n = 2$ to $n = 3$ are sometimes not very easy and cited Fermat's Last Theorem as an illustration.

Even in the three-period BDT model, which we will discuss in the following section, the most efficient algorithm for QE is already difficult to handle. In subsection 3.2, we give an efficient algorithm for our problem in the three-period BDT model using ideas from QE.

2.1 *Calibration of the modified BDT model*

In this subsection, we briefly discuss a modified version of the BDT model which takes the term structure of short rate volatilities as input instead of the term structure of yield volatilities. In this case, the calibration is a lot simpler. First, the term structure of short rate volatilities can be inferred from the market prices of interest rate caps. Second, the calibration reduces to solving one non-linear equation since under the assumption of lognormality of the short rates, we have

$$r(n,i) \;=\; r(n,i-1)e^{2\sigma_r(n)\sqrt{\Delta t}}, \quad \text{for} \quad i = 1,2,\ldots,n$$

where $\hat{\sigma}_r(n)$ is the input short rate volatility for the nth period. In other words, the parameter β_n in (5) becomes known and is equivalent to $e^{2\hat{\sigma}_r(n)\sqrt{\Delta t}}$. Hence there is only one equation with one unknown for this calibration.

For the modified BDT model, Theorem 2.1 of Sandmann and Sondermann (1993) provides a necessary and sufficient condition under which it is possible to calibrate a BDT model as long as the short rate volatilities are positive and finite. They show that there exists a BDT model (with positive short rates) if and only if the forward rates are positive. This result can easily be shown as follows:

Suppose the modified BDT model has been calibrated up to period n. To proceed one more period, we need to solve the following equation (from (5))

$$\alpha_{n+1} \;=\; \sum_{i=0}^{n} \frac{A(n,i)}{1 + r(n,0)\beta_n^i \Delta t} \tag{16}$$

where $\beta_n = e^{2\hat{\sigma}_r(n)\sqrt{\Delta t}}$. Let f_n be the one-period forward rate from period n to $n+1$. Then we have

$$\frac{1}{1 + f_n \Delta t} \;=\; \frac{\alpha_{n+1}}{\alpha_n} \;=\; \sum_{i=0}^{n} \frac{a_i}{1 + r(n,0)\beta_n^i \Delta t}$$

where

$$a_i \;=\; \frac{A(n,i)}{\alpha_n} \geq 0$$

and $\Sigma_i\, a_i = 1$ since $\Sigma_i\, A(n,\, i) = \alpha_n$.

If $r(n,\, 0)\ \beta_n^i$ are positive for all $0 \leq i \leq n$,
then

$$\frac{1}{1 + r(n,0)\beta_n^i \Delta t} \in (0,1)$$

This implies that the convex combination

$$\sum\nolimits_{i=0}^{n} \frac{a_i}{1 + r(n,0)\beta_n^i \Delta t} \in (0,1)$$

Consequently, we must have $f_n > 0$.

Conversely, suppose $f_n > 0$. We need to show that there exists a unique positive number $r(n, 0)$ such that (16) holds. Let

$$h(x) = \sum_{i=0}^{n} \frac{a_i}{1 + x\beta_n^i \Delta t} - \frac{1}{1 + f_n \Delta t}$$

First note that $h(x)$ is strictly decreasing. Second, by assumption we have

$$h(0) = 1 - \frac{1}{1 + f_n \Delta t} > 0$$

Third,

$$h(+\infty) = -\frac{1}{1 + f_n \Delta t} < 0$$

This implies that there exists a unique positive root for $h(x) = 0$, as required.

3. Three-period BDT model

3.1 *A sufficient condition*

In this section, we consider the calibration issue for the original three-period BDT model. We assume that the first two periods have already been calibrated successfully so that we only need to find $r(2, 0)$, $r(2, 1)$, and $r(2, 2)$ such that the resulting interest rate lattice matches to the input term structures $\hat{Y}(3)$ and $\hat{\sigma}_Y(3)$. Although this is only the third period calibration, we show that the BDT lattice may not exist for certain term structures. To examine these conditions, we first note that eliminating v in (6) and (10) with $n = 2$, the parameter u becomes the root of a polynomial equation of degree 4; i.e. $g_1(u) = 0$ where

$$g_1(x) = x^4 + \frac{2}{\eta_3 - 1} x^3 + \left[\frac{\eta_3^2 + 1}{(\eta_3 - 1)^2} - \frac{2\alpha_3}{\alpha_1} \right] x^2 - \frac{4\alpha_3}{\alpha_1(\eta_3 - 1)} x - \frac{2\alpha_3}{\alpha_1(\eta_3 - 1)^2} \tag{17}$$

The first result can be stated as follows:

Theorem 1 If there exist positive short rates in the interval $(0, \omega)$, then

$$g_1(\delta(\omega)) > 0,$$

where

$$\delta(x) = \min\left\{ \sqrt{\frac{1}{1 + r(1,1)\Delta t}}, \sqrt{\frac{2\alpha_3}{\alpha_1} - \frac{1}{[1 + r(1,0)\Delta t](1 + x\Delta t)}} \right\}.$$

In other words, if $g_1(\delta(\omega)) \leq 0$, then there exists no positive short rates in the interval $(0, \omega)$.

Proof: See Appendix B.

A consequence of the above theorem is that calibration of the BDT model is not feasible when the yield curve is increasing sharply while the yield volatility curve is decreasing dramatically. This observation can be verified as follows:

Substituting $\delta(\omega)$ into (17), we obtain

$$g_1(\delta(\omega)) = \left[\delta(\omega)^4 - \frac{2\alpha_3}{\alpha_1}\delta(\omega)^2\right] + \frac{1}{\eta_3 - 1}\left[2\delta(\omega)^3 - \frac{4\alpha_3}{\alpha_1}\delta(\omega)\right] + \frac{1}{(\eta_3 - 1)^2}\left[(\eta_3^2 + 1)\delta(\omega)^2 - \frac{2\alpha_3}{\alpha_1}\right]$$

Assuming that the yield volatility is positive, we have $\eta_3 > 1$. Notice that the terms in the first two square brackets are always negative. This implies that $g_1(\delta(\omega)) \leq 0$ if the term in the third bracket is also negative; i.e.

$$(\eta^2 + 1)\delta(\omega)^2 < \frac{2\alpha_3}{\alpha_1}$$

Hence we have the following corollary:

Corollary 2

(a) Suppose

$$\delta(\omega) = \sqrt{\frac{2\alpha_3}{\alpha_1} - \frac{1}{[1 + r(1,0)\Delta t](1 + \omega\Delta t)}}$$

There exists no short rates $\{r(2, 0), r(2, 1), r(2, 2)\}$ in the range $(0, \omega)$ if

$$\sigma_Y(3) \leq -\frac{1}{4\sqrt{\Delta t}}\log\left(\frac{2\alpha_3}{\alpha_1}[1 + r(1,0)\Delta t](1 + \omega\Delta t) - 1\right) \tag{18}$$

and

$$\alpha_1 < 2\alpha_3[1 + r(1,0)\Delta t](1 + \omega\Delta t)$$

(b) Suppose

$$\delta(\omega) = \sqrt{\frac{1}{1 + r(1,1)\Delta t}}$$

There exists no short rates $\{r(2, 0), r(2, 1), r(2, 2)\}$ in the range $(0, \omega)$ if

$$\sigma_Y(3) \leq \frac{1}{4\sqrt{\Delta t}}\log\left(\frac{2\alpha_3}{\alpha_1}[1 + r(1,1)\Delta t] - 1\right) \tag{19}$$

and

$$\alpha_1 < 2\alpha_3[1 + r(1,1)\Delta t]$$

3.2 *Necessary and sufficient conditions*

Theorem 1 established a sufficient condition for the third period calibration to be feasible. In this subsection, we provide a necessary and sufficient condition for the short rates to lie in the interval $(0, \omega)$, where $\omega > 0$.

We assume the largest short rate lies on the top branch of the BDT lattice and denote it as y^*, then the other two short rates in the third period are y^*/β_3, y^*/β_3^2, where $\beta_3 \geq 1$. Since y^* is the largest attainable rate, it is sufficient to consider the conditions on y^* for which it lies in $(0, \omega)$.

From (7) and (8), u and v must satisfy the following equations:

$$\frac{1}{1+y} + \frac{1}{1+y/\beta} = 2(1+c)u^2 \tag{20}$$

$$\frac{1}{1+y/\beta} + \frac{1}{1+y/\beta^2} = 2(1+b)v^2 \tag{21}$$

where $\beta = \beta_3$, $y = y^* \Delta t$, $b = r(1, 0)\Delta t$ and $c = r(1, 1)\Delta t$. The assumption that $\beta \geq 1$ implies that the bound on v is

$$B_1 \leq v \leq B_2 \tag{22}$$

where

$$B_1 = \sqrt{\frac{2\alpha_3}{\alpha_1}\frac{1+c}{2+b+c}} \quad \text{and} \quad B_2 = \sqrt{\frac{2\alpha_3}{\alpha_1} - \frac{1}{(1+c)(1+\omega)}}$$

Define m and n as

$$m = \frac{1}{2(1+c)u^2} = \frac{1}{2(1+c)(2\frac{\alpha_3}{\alpha_1} - v^2)} \tag{23}$$

$$n = \frac{1}{2(1+b)v^2} \tag{24}$$

Note that $m \geq n$ since $\beta \geq 1$. Substituting the above expressions m and n into (20) and (21), we obtain

$$y^2 + (\beta+1)(1-m)y + (1-2m)\beta = 0 \tag{25}$$

$$y^2 + \beta(\beta+1)(1-n)y + (1-2n)\beta^3 = 0 \tag{26}$$

Eliminating the y^2 term gives

$$y = \frac{\beta[2m - 1 - (2n-1)\beta^2]}{(\beta+1)[(n-1)\beta - (m-1)]}$$

or

$$y^* = \frac{\beta[2m - 1 - (2n-1)\beta^2]}{\Delta t(\beta+1)[(n-1)\beta - (m-1)]} \tag{27}$$

To ensure that the condition $0 \leq y^* \leq \omega$ is satisfied, we need to consider the cases where the denominator in (27) is either positive or negative.

Case 1: $(n-1)\beta - (m-1) > 0$

In this situation, the only admissible case is $n > 1$. Here is why this is the only case.

If $n = 1$ then $-(m-1) > 0$ which implies that $m < 1$ contradicting the assumption that $m \geq n$.

If $n < 1$, then $\beta < (1 - m)/(1 - n) < 1$ because $(m \geq n)$ and this contradicts the assumption that $\beta \geq 1$. For $n > 1$, the conditions on β are

$$\beta > \frac{m - 1}{n - 1} \tag{28}$$

$$\beta \leq \sqrt{\frac{2m - 1}{2n - 1}} \tag{29}$$

and

$$\omega \Delta t(\beta + 1)[(n - 1)\beta - (m - 1)] - \beta[2m - 1 - (2n - 1)\beta^2] > 0 \tag{30}$$

Let Ψ_1 be the expression on the left-hand side of the above inequality with m and n replaced by (23) and (24). Ψ_1 becomes a function in terms of the unknown variables v and β. An equivalent condition to (30) becomes

$$\Psi_1(\beta, \nu) > 0 \tag{31}$$

where v is a root (satisfying the boundary condition (22)) of the function g_2 defined as

$$g_2(x) = x^4 + \frac{2\eta_3}{\eta_3 - 1}x^3 + \left[\frac{\eta_3^2 + 1}{(\eta_3 - 1)^2} - \frac{2\alpha_3}{\alpha_1}\right]x^2 + \frac{4\alpha_3\eta_3}{\alpha_1(\eta_3 - 1)}x - \frac{2\alpha_3\eta_3}{\alpha_1(\eta_3 - 1)^2} \tag{32}$$

The above function is derived from (6) and (10) by eliminating u.[5] Inequality (31) provides one condition for which β and v must be jointly satisfied.

In a similar manner, we define

$$\Psi_2 = \beta - \frac{m - 1}{n - 1}$$

and

$$\Psi_3 = \beta^2 - \frac{2m - 1}{2n - 1}$$

and substituting (23) and (24) into the above two expressions. Inequalities (28) and (29) are respectively, equivalent to

$$\Psi_2(\beta, \upsilon) > 0 \tag{33}$$

$$\Psi_3(\beta, \upsilon) < 0 \tag{34}$$

For a given root v, there may exist many possible values of β for which conditions (31), (33), and (34) are fulfilled. Hence we need another condition on β so that the uniqueness of β is ensured. This is achieved by substituting (27) into (25). If we denote the resulting expression by $\Psi_4(\beta, v)$, β is then computed from the following equation

$$\Psi_4(\beta, \upsilon) = 0 \tag{35}$$

The above series of steps provides necessary conditions for which the short rates lie in the required range $(0, \omega)$. It remains to consider the other situation where the denominator in (27) is negative. In

[5] Alternatively, $g_2(x)$ can be obtained from $g_1(x)$ by replacing η_3 in (17) by $1/\eta_3$.

this case, there are two admissible possibilities depending on the value n. We summarize the results as follows:

Case 2a: $(n-1)\beta-(m-1)<0$ and $n>1$
Using the above notation, the conditions on v and β are

$$\Psi_1(\beta,\upsilon) < 0$$
$$\Psi_2(\beta,\upsilon) < 0$$
$$\Psi_3(\beta,\upsilon) \geq 0$$
$$\Psi_4(\beta,\upsilon) = 0$$

Case 2b: $(n-1)\beta-(m-1)<0$ and $n\leq 1$
In this case, the conditions on v and β are

$$\Psi_1(\beta,\upsilon) < 0$$
$$\Psi_2(\beta,\upsilon) > 0$$
$$\Psi_3(\beta,\upsilon) \geq 0$$
$$\Psi_4(\beta,\upsilon) = 0$$

From the above discussion, we also know that the process can be reversed. The reason is as follows: the inequality $B_1 \leq v \leq B_2$ implies that $m \geq n$. Then in case 1, both relations $\Psi_2(\beta, v) > 0$ and $\Psi_3(\beta, v) < 0$ imply that $y > 0$, and $\Psi_1(\beta, v) > 0$ yields $y^* < \omega$. Then both relations $\Psi_4(\beta, v) = 0$ and $g_2(v) = 0$ are equivalent to the original non-linear relations (5) and (6). The results for Cases 2a and 2b are similar.

To conclude this section, we provide an algorithm which checks the existence of the third period BDT short rates in $(0, \omega)$.

Step 1. First check whether the equation $g_2(v) = 0$ has a solution in the range (B_1, B_2). This can be accomplished using Sturm's algorithm (see Appendix C for a brief description). Alternatively, some mathematical software packages such as Maple have a built-in version of Sturm's algorithm. If no such root v exists, stop.

Step 2. If there exists such roots, there are at most four roots. Since the degree of g_2 is 4, these roots can be found relatively easily. For each v, we obtain the corresponding β by solving $\Psi_4(\beta, v) = 0$. Sturm's algorithm can again be used as a first step to check the existence of the root. If the root β exists, then go to step 3; otherwise stop.

Step 3. For each root β (only finitely many) and v, compute m and n using (23) and (24) and check if any of the following conditions holds:

$$[3.1].\ \Psi_1(\beta,\upsilon) > 0, \Psi_2(\beta,\upsilon) > 0, \Psi_3(\beta,\upsilon) < 0 \text{ and } \upsilon \leq \sqrt{\frac{1}{2(1+b)}}.^{6}$$

[6] The last condition on v is equivalent to the condition $n \geq 1$.

[3.2a]. $\Psi_1(\beta, \upsilon) < 0, \Psi_2(\beta, \upsilon) < 0, \Psi_3(\beta, \upsilon) \geq 0$ and $\upsilon \leq \sqrt{\frac{1}{2(1+b)}}$.

[3.2b]. $\Psi_1(\beta, \upsilon) < 0, \Psi_2(\beta, \upsilon) > 0, \Psi_3(\beta, \upsilon) \geq 0$ and $\upsilon \geq \sqrt{\frac{1}{2(1+b)}}$.

If one of the above conditions holds, then there exists short rates, say y^*, y^*/β, y^*/β^2, in the range $(0, \omega)$, where y^* can be determined from (27). Otherwise, positive short rates in the range $(0, \omega)$ do not exist.

4. General case

In the previous section, we provided a necessary and sufficient condition for the existence of the positive short rates for the three-period BDT model. We saw that even with three periods, the task was already very challenging. Hence it seems virtually impossible to extend the methodology to the general n-period case.

In this section we assume for a given input term structures, the model has been calibrated (successfully) up to n periods. We then provide a sufficient condition that jointly characterizes $\hat{Y}(n+1)$ and $\hat{\sigma}_Y(n+1)$ for which the calibration would fail in the $(n+1)$th period. More precisely, we summarize the result in the following theorem.

Theorem 3 Let $a = [P_d(n)]^{\frac{1}{n}}$, $b = [P_u(n)]^{\frac{1}{n}}$, and

$$\eta(x) = \frac{x}{1-x}\left[\left(2\frac{\alpha_{n+1}}{\alpha_1} - x^n\right)^{-\frac{1}{n}} - 1\right] \tag{36}$$

for $x > 0$. Assume for any arbitrary term structure, the BDT model has been calibrated up to n periods. In the calibration to the $(n+1)$th period, there exists no positive short rates if the inputs $\hat{Y}(n+1)$ and $\hat{\sigma}_Y(n+1)$ satisfy any of the following conditions:

$$\text{(i)} \quad \hat{\sigma}_y(n+1) > \frac{1}{2\sqrt{\Delta t}} \log(\eta(a)) \tag{37}$$

$$\text{(ii)} \quad \hat{\sigma}_y(n+1) < \frac{1}{2\sqrt{\Delta t}} \log\left(\frac{1}{\eta(b)}\right) \tag{38}$$

Proof: See Appendix D.

Remark 1. The proof to the above theorem only requires the monotone-property of the two-variable function $f(x, y)$ and the basic relation among the Arrow–Debreu securities. The same conclusion applies to any interest rate tree as long as the short rates satisfy the BDT relations; i.e. $r(n, i) = i(n, 0)\beta_n^i$, $\forall n, i$.

Remark 2. If $\eta(a)\eta(b) \leq 1$, then there exists no positive short rates in the $(n+1)$th period regardless of the value $\hat{\sigma}_y(n+1)$.

Table 1. Examples of input term structures.

Example	(*a*)	(*b*)	(*c*)	(*d*)	(*e*)	(*f*)
N	4	12	52	52	10	10
Δt (yrs)	1/4	1/12	1/52	1/12	1/4	1/12
$T = N\,\Delta t$ (yrs)	1	1	1	$4\frac{1}{3}$	$2\frac{1}{2}$	$\frac{5}{6}$
$\hat{Y}$ (1)	8%	8%	8%	8%	10%	10%
$\hat{Y}$ (N)	4%	4%	4%	4%	5%	5%
$\hat{\sigma}_Y$ (1)	20%	20%	20%	20%	20%	20%
$\hat{\sigma}_Y$ (N)	20%	20%	20%	20%	17%	17%

5. Numerical examples

In this section, we present numerical examples which illustrate the calibration issues based on the results established in the earlier sections.

Table 1 gives details of the cases which we used to construct N-period BDT models. In particular, the first three examples, (*a*), (*b*), and (*c*), have identical term structures in the sense that all three of them have a flat yield volatility of 20% while the yield curve is decreasing linearly from 8% to 4% over one year horizon. The only difference is the number of time steps fitted to these term structures. Example (*a*) uses 4 periods, (*b*) uses 12 periods, while (*c*) uses 52 periods. This accounts for the difference in magnitude of Δt reported in the third row of the table. In each of these examples, we did not encounter any problem in constructing the BDT models (for varying N).

Suppose we make a minor change to the input yield volatilities in these examples (*a*), (*b*), and (*c*) and attempt to re-calibrate the BDT lattice. The minor change is by perturbing the N-period yield volatility from 20% to 19% (i.e. $\hat{\sigma}_Y(N) = 19\%$). In all three situations, we found that we can only calibrate up to $N-1$ periods and fail in the Nth period. It might be argued that this phenomenon results from to the numerical methods used to solve the two non-linear equations (5) and (6), for instance, if a Newton–Raphson procedure were used to solve these equations. Failure to find convergence does not necessary imply that there exists no solution. This could merely be due to a poor set of bad initial values in carrying out the iteration process. It is well-known that the convergence of the Newton–Raphson algorithm crucially relies on the initial values.

We now show that Theorem 3 can be used to explain the failure to calibrate the nth period short rates. In $N-1$ periods, the quantities $P_d(N-1)$ and $P_u(N-1)$ are readily available. From the assumed value of the N-period yield (i.e. $\hat{Y}(N)$), we can determine which condition(s) in Theorem 3 is(are) satisfied and hence find the appropriate bound of $\hat{\sigma}_Y(N)$ for which positive short rates in the Nth period are not possible. It turns out that in all three cases, either condition (*i*) or condition (*ii*) is satisfied. The ranges on the yield volatilities are reported in Table 2. For instance for case (*a*), there does not exist positive short rates in the fourth period when the yield volatility is either greater than 20.568% or less than 19.238%. In our modified examples, we have $\hat{\sigma}_Y(4) = 19\%$, which falls in the required range. This explains the failure in solving the fourth period short rates that calibrate to this yield volatility. It should be emphasized that the above result is based on our theory and is independent of the numerical techniques.

Table 2. The implied ranges on $\hat{\sigma}_Y(N)$ from Theorem 3 for examples in Table 1.

Example	(*a*)	(*b*)	(*c*)	(*d*)	(*e*)	(*f*)
N	4	12	52	52	10	10
(*i*): $\hat{\sigma}_Y(N) >$	20.568%	20.057%	20.005%	20.011%	17.481%	17.425%
(*ii*): $\hat{\sigma}_Y(N) <$	n/a	n/a	n/a	n/a	n/a	n/a
(*ii*): $\hat{\sigma}_Y(N) <$	19.238%	19.930%	19.994%	19.999%	17.135%	17.224%
(*i*): $\hat{\sigma}_Y(N) >$	n/a	n/a	n/a	n/a	n/a	n/a

The results in Table 2 also indicate that as the size of the period becomes smaller, the convergence of the BDT model becomes more sensitive. For example, a small perturbation of the n-period yield volatility, say to 19.9%, would lead to no positive short rates for Example (*c*) with $\Delta t = \frac{1}{52}$ while the result is inconclusive for both (*b*) and (*c*) with larger time steps.

We now consider the rest of the examples ((*d*), (*e*) and (*f*)) in Table 1. Example (*d*) is similar to (*c*) except that each period is of length $\frac{1}{12}$ years so that the total time horizon in 52 periods are $4\frac{1}{3}$ years. In example (*e*), the yield curve is downward sloping, decreasing linearly from 10% to 5% over $2\frac{1}{2}$ years. The yield volatility is also downward sloping which decreases linearly from 20% to 17% over the same time horizon. The input term structures for example (*f*) also exhibits the same shape as (*e*) except that the curves are spread over a much shorter horizon; i.e. $\frac{5}{6}$ years. We will use a 10-period BDT lattice to fit the term structures given in examples (*e*) and (*f*).

It is found that even with a much longer time horizon (compared to (*c*)), there is no problem in calibrating the term structure in (*d*). On the other hand, the calibration fails in the tenth period for the decreasing term structures in (*e*) and (*f*). This issue can again be addressed using Theorem 3. The last three columns in Table 2 summarize the results. Theorem 3 indicates that if the 10-period yield volatility in cases (*e*) and (*f*) is less than 17.14% and 17.22%, respectively, then there exists no positive short rates in this period. The input term structure, which is $\hat{\sigma}_Y(10) = 17\%$, clearly falls within the range and hence accounts for the break-down of the BDT in the tenth period.

An alternative characterization of Theorem 3 is to examine the 'regions of no solution'. The conditions in Theorem 3 jointly provide the bounds on $\hat{Y}(n)$, and $\hat{\sigma}_Y(n)$ for which the BDT model cannot be calibrated. Figure 1 depicts the regions of interest for Example (*e*). The shaded region is derived from condition (*i*) while the striped region from condition (*ii*). Hence any input data $(\hat{Y}(n), \hat{\sigma}_Y(n))$ which lie in these zones makes the calibration infeasible. This also explains the phenomenon observed in Example (*e*) where the input data lies within the failure zone (marked with an X on the figure).

It should be pointed out that in all the examples we have considered, the forward rates are *positive*. If the n-period yield volatility satisfies the bounds in Table 2, then positive short rates do not exist. This indicates that the implied short rate volatility is either negative or infinite, a consequence of Theorem 2.1 of Sandmann and Sondermann (1993) (or see Section 2). For instance, in Example (*b*) with $\hat{Y}(12) = 4\%$ and $\hat{\sigma}_Y(12) = 20\%$, the implied short rate volatility turns out to be 36.65%. Theorem 2.1 of Sandmann and Sondermann ensures that the BDT model can be calibrated since the forward rates are positive and the short rate volatility is positive and finite. Now if one changes the magnitude of $\hat{\sigma}_Y(12)$ so that it approaches the bounds in Table 2, the implied short rate volatility must necessarily be approaching zero or infinity. This is illustrated in Table 3 which gives the implied short rate volatility in the last period by fitting to $\hat{\sigma}_Y(12) = 19.994\%$, 19.995%, 20.005% and 20.006%

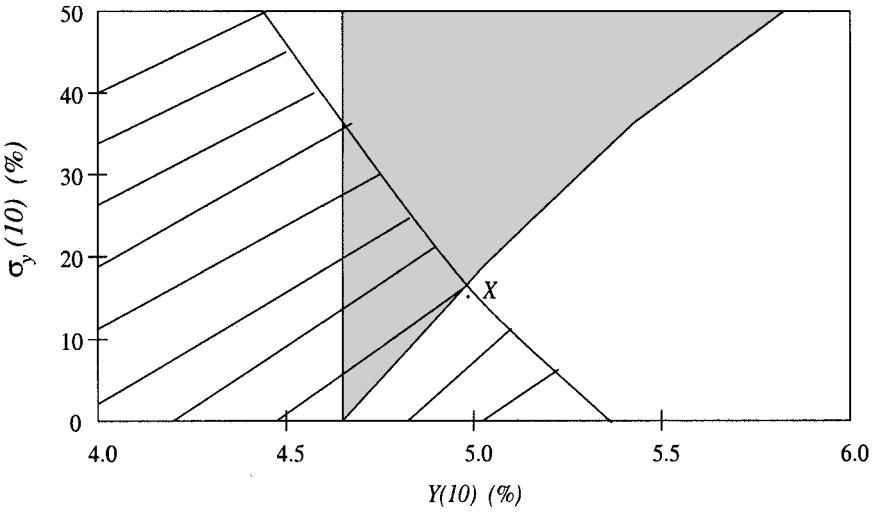

Fig. 1. Regions of no solution for Example (*e*).

Table 3. The implied short rate volatility in Example (*b*) for different input $\hat{\sigma}_Y$ (*n*).

$\hat{\sigma}_Y$ (12) (%)	*n-period implied short rate volatility (%)*
19.994	3.554
19.995	9.050
20.005	64.712
20.006	70.414

while maintaining $\hat{Y}$ (12) = 4%. In these cases, even with a small perturbation of the yield volatility, the implied short rate volatility changes dramatically. More importantly, when the yield volatility is approaching the bound given by condition (*i*), the implied short rate volatility gets larger. On the other hand, when the yield volatility is approaching the bound by condition (*ii*), the implied short rate volatility gets smaller.

6. Conclusion

We already know that all one-factor models such as the BDT model have limitations in modelling yield curve behaviour. Litterman and Scheinkman (1991) find that three factors are required to provide an adequate representation of yield curve dynamics. Hull and White (1995) have noted that if we overfit a one-factor Markovian model, we end up with unrealistic dynamics for the future evolution of volatility. Indeed they suggest that one should only fit the initial term structure of bond prices and

not overparametrize the model. Radhakrishnan (1998) points out that the BDT model generates pricing errors irrespective of whether the model is fitted to the yield volatilities or the short rate volatilities. By using the HJM model as a benchmark, he finds that if the yield volatility is used to calibrate the corresponding BDT model, the BDT model underprices options with long maturities. Conversely, he finds that if the short rate volatility is used, the BDT model overprices options on long term bonds.

This paper has pointed out some additional technical problems that may arise when fitting a BDT lattice to an input set of yields and term structure of yield volatilities. The situation is more complicated when we fit the yield volatility rather than the short rate volatility. We find explicit mathematical conditions which indicate when it is feasible to fit a BDT model. In this case they involve joint restrictions on both the input bond prices and the yield volatilities. When we use the yield volatilities as input, the resulting mathematical conditions to ensure a feasible calibration are quite complicated. If we use the short rate volatilities as the input, then both the economic intuition and the mathematical restrictions are much simpler. Analysis confirms that the conventional practice of fitting the model to the short rate volatilities rather than the yield volatilities has theoretical as well as practical advantages.

References

Arnon, D. (1981) Algorithms for the Geometry of Semi-Algebraic Sets. PhD thesis, Technical Report no. 436, Computer Science Dept. University of Wisconsin-Madison.

Black, F., Derman, E. and Toy, W. (1990) A one-factor model of interest rates and its application to treasury bond Options, *Financial Analysts Journal*, **46**(1), 33–9.

Brennan, M.J. and Schwartz, E. (1979) A continuous-time approach to pricing of bonds, *Journal of Banking and Finance*, **3**, 135–55.

Cox, J.C., Ingersoll, J.E. and Ross, S.A. (1985) A theory of the term structure of interest rates, *Econometrica*, **53**, 385–407.

Collins, G.E. (1975) Quantifier elimination for real closed fields by cylindrical algebraic decomposition. In G. Goos and J. Hartmanis, editors, *Automata Theory and Formal Languages 2nd Gl Conference, Springer Lecture Notes in Computer Science*, Vol. 33, pp. 134–83.

Heath, D., Jarrow, R. and Morton, A. (1990) Bond pricing and the term structure of interest rates: a discrete time approximation, *Journal of Financial and Quantitative Analysis*, **25**(4), 419–40.

Ho, T.S.Y. and Lee, S. (1986) Term structure movements and pricing interest rate contingent claims, *Journal of Finance*, **41**(5), 1011–29.

Hull, J. and White, A. (1995) A note of the models of Hull and White for pricing options on the term structure: a response, *Journal of Fixed Income*, **5**(2), 97–102.

Hull, J. and White, A. (1993) One-factor interest rate models and the valuation of interest rate derivative securities, *Journal of Financial and Quantitative Analysis*, **28**, 225–254.

Jacobson, N. (1985) *Basic Algebra*, Volume I (second edition). W.H. Freeman and Company, New York.

Jamshidian, F. (1991) Forward induction and construction of yield curve diffusion models, *Journal of Fixed Income*, **June**, 62–74.

Litterman, R. and Scheinkman, J. (1991) Common factors affecting bond returns, *Journal of Fixed Income*, **1**, 54–61.

Pan, V.Y. (1997) Solving a polynomial equation: some history and recent progress, *SIAM Review*, **30**(2), 187–220.

Radhakrishnan, A.R. (1998) Mispricing of discount bond options in the Black–Derman–Toy model calibrated to term structure and cap volatilities: an empirical study. Working paper, New York University, August.

Sandmann, K. and Sondermann, D. (1993) A term structure model and the pricing of interest rate derivatives, *The Review of Futures Markets*, **12**(2), 391–423.

Tarski, A. (1951) *A Decision Method for Elementary Algebra and Geometry*, second edition. University of California Press, Berkeley.

Vasicek, O.A. (1977) An equilibrium characterization of the term structure, *Journal of Financial Economics*, **5**, 177–88.

Appendix A: Quantifier elimination

We briefly review the main theorem of Quantifier Elimination (QE) in this appendix.

Theorem 4 (QE Theorem) Let $A = \Re[t_1, \ldots, t_r]$, $B = A[x_1, \ldots, x_n]$ where the t's and x's are indeterminates and $\Re$ is the real number field. Let

$$\Gamma = \{F_1, \ldots, F_m, G\} \subseteq B$$

Then we can determine in a finite number of steps (and in a constructive way) a finite collection $\{\Gamma_1, \ldots, \Gamma_s\}$ where

$$\Gamma_j = \{f_{j,1}, \ldots, F_{j,m_j}, g_j\} \subseteq A$$

such that for any $(c_1, \ldots, c_r) \in \Re^{(r)}$ the system of equations and equation $\Gamma(c_1, \ldots, c_r)$:

$$\begin{aligned} F_1(c_1, \ldots, c_r; x_1, \ldots, x_n) &= 0 \\ &\vdots \\ F_m(c_1, \ldots, c_r; x_1, \ldots, x_n) &= 0 \\ G(c_1, \ldots, c_r; x_1, \ldots, x_n) &\neq 0 \end{aligned}$$

is solvable for the xs in R if and only if the c_i satisfy one of the systems $\Gamma_j(c_1, \ldots, c_r)$:

$$\begin{aligned} f_{j,1}(c_1, \ldots, c_r) &= 0 \\ &\vdots \\ f_{j,m_j}(c_1, \Lambda, c_r) &= 0 \\ g_j(c_1, \ldots, c_r) &\neq 0 \end{aligned}$$

for every $1 \leq j \leq s$. There are no variables $x_1, \ldots, x_n$ in the functions $f_{j,k}$, g_j, $K\ 1 \leq j \leq s$, $K\ 1 \leq k \leq m_j$.

This elimination process was first given by Tarski (1951), and the techniques traced back to the work of Sturm, Euler and Bezout in the 18th century. Tarski's original method was not practical and was substantially modified and improved by Collins (1975). Collins's algorithm has been implemented in a computer algebra system by Arnon (1981). Even in the three-step case, this efficient algorithm is so complicated that it is impossible to give the criterion explicitly by the analysis of the computing time given in Collins's original paper. Nevertheless, we can still provide criteria to check whether there exists reasonable interest rates in the BDT model as discussed in Section 4.

Appendix B: Proof of Theorem 1

First we need the following lemma:

Lemma 5 If

$$r(2,0), r(2,1), r(2,2) \in (0, \omega) \tag{39}$$

then

$$u^2 \in (\max[L_1, L_2], \min[M_1, M_2]) \tag{40}$$

where

$$\begin{aligned} L_1 &= \frac{1}{(1 + \omega\Delta t)[1 + r(1,1)\Delta t]} \\ L_2 &= \frac{2\alpha_3}{\alpha_1} - \frac{1}{1 + r(1,0)\Delta t} \\ M_1 &= \frac{1}{1 + r(1,1)\Delta t} \\ M_2 &= \frac{2\alpha_3}{\alpha_1} - \frac{1}{[1 + r(1,0)\Delta t](1 + \omega\Delta t)} \end{aligned}$$

Proof. Define $r_0 = (1 + r(2, 0)\Delta t)^{-1}$, $r_1 = (1 + r(2, 1)\ \Delta t)^{-1}$, $r_2 = (1 + r(2, 2)\ \Delta t)^{-1}$. It follows from (39) that

$$r_0, r_1, r_2 \in \left(\frac{1}{1 + \omega\Delta t}, 1\right)$$

By definition, the variable u^2 and v^2 can be expressed as

$$u^2 = \frac{1}{2[1 + r(1,1)\Delta t]}(r_1 + r_2) \tag{41}$$

$$v^2 = \frac{1}{2[1 + r(1,0)\Delta t]}(r_0 + r_1) \tag{42}$$

Substituting (42) into (10) and rearranging, we obtain

$$u^2 = \frac{1}{2}\left(\frac{4\alpha_3}{\alpha_1} - \frac{r_0 + r_1}{1 + r(1,0)\Delta t}\right) \tag{43}$$

Since

$$r_0 + r_1, r_1 + r_2 \in \left(\frac{2}{1 + \omega\Delta t}, 2\right)$$

Equations 41 and 43 imply that the bounds on u^2 are (L_1, M_1) and (L_2, M_2), respectively. This completes the proof to the lemma.

To prove Theorem 1. First note that the root of g_1 solves (6) and (10). Furthermore, the positivity of the short rates implies that we are only interested in the positive root of g_1. It follows from the above lemma that condition (39) implies that $0 < u < \delta(\omega)$.

Suppose $g_1(\delta(\omega)) \leq 0$, we need to prove $g_1(x)$ has no root in $(0, \delta(\omega))$. Let e_j denote the coefficient of x^j in (17). We have $g_1(0) = -e_0 < 0$ and $g_1'(0) = -e_1 < 0$. This implies that we only need to examine the behaviour of $g_1(x)$ in $(0, \delta(\omega))$. We argue that the function g_1 is either decreasing from $x = 0$ and then increasing until $x = \delta(\omega)$, or is decreasing in the entire range. This is equivalent to saying that $g_1'(x) = 0$ has at most one root in $(0, \delta(\omega))$

Suppose α is the root of $g_1'(x) = 0$ in the interval $(0, \delta(\omega))$, and let β, γ be other roots of this equation.[7] Then we have

$$\alpha + \beta + \gamma = -\frac{3e_3}{4}$$

$$\alpha\beta + \alpha\gamma + \beta\gamma = \frac{e_2}{2}$$

$$\alpha\beta\gamma = e_1$$

If one of β, γ is not a real number, so is the other. In this case the proof is complete. Suppose both β, γ are real numbers. From the last equation we have $\beta\gamma > 0$. Therefore, either $\beta > 0, \gamma > 0$ or $\beta < 0, \gamma < 0$. However, $\beta + \gamma = -3e_1/4 - \alpha < 0$, which leads to $\beta < 0, \gamma < 0$. Hence, $g_1'(x)$ has at most one root in the interval $(0, \delta(\omega))$. This completes the proof of Theorem 1.

Appendix C: Sturm's Theorem

In this appendix, we summarize the key result of Sturm's Theorem which is taken from Jacobson (1985). If $c = \{c_1, c_2, \ldots, c_m\}$ is a finite sequence of non-zero elements of $\Re$, then we define the *number of variations in sign of c* to be the number of i, $1 \leq i \leq m-1$, such that $c_i c_{i+1} < 0$. On the other hand, if c is an arbitrary sequence of elements in $\Re$, then the number of variations in sign of c is defined to be the number of variations in the sign of the subsequence c' obtained by dropping the 0s in c.

Now let $f(x)$ be any polynomial in $\Re[x]$ of positive degree. We define the *standard sequence* for $f(x)$ as

$$\begin{aligned}
f_0(x) &= f(x) \\
f_1(x) &= f'(x) \qquad\qquad \text{(formal derivative of } f(x)\text{)} \\
f_0(x) &= q_1(x)f_1(x) - f_2(x), \quad \deg f_2 < \deg f_1 \\
&\vdots \qquad\qquad\qquad\qquad \vdots \\
f_{i-1}(x) &= q_i(x)f_i(x) - f_{i+1}(x), \quad \deg f_{i+1} < \deg f_i \\
&\vdots \qquad\qquad\qquad\qquad \vdots \\
f_{s-1}(x) &= q_s(x)f_s(x) \qquad\qquad \text{i.e. } f_{s+1}(x) = 0
\end{aligned} \tag{44}$$

[7] Note that β, γ may be complex numbers.

Thus, the $f_i(x)$ are obtained by modifying the Euclid algorithm for finding the g.c.d. of $f(x)$ and $f'(x)$ in such a way that the last polynomial obtained at each stage is the negative of the remainder in the division process.

We now state Sturm's Theorem.

Theorem 6 (Sturm's Theorem) Let $f(x)$ be a polynomial of positive degree with coefficients in the real number field $\Re$ and let $\{f_0(x) = f(x), f_1(x) = f'(x), \ldots, f_s(x)\}$ be the standard sequence (44) for $f(x)$. Assume $[a, b]$ is an interval such that $f(a) \neq 0, f(b) \neq 0$. Then the number of distinct roots of $f(x)$ in (a, b) is $V_a - V_b$ where V_c denotes the number of variations in sign of $\{f_0(c), f_1(c), \ldots, f_s(c)\}$.

Remark: The Newton–Raphson algorithm is a commonly used technique for finding the zeros of the polynomials. There are many other efficient algorithms for approximating polynomial zeros as well. See the recent survey paper by Pan (1997).

Appendix D: Proof of Theorem 3

Proof. Substituting (6) into (10) and eliminating u, we obtain

$$v^n + \left(1 - \eta_{n+1} + \frac{\eta_{n+1}}{v}\right)^{-n} - 2\frac{\alpha_{n+1}}{\alpha_1} = 0 \tag{45}$$

or equivalently,

$$\frac{v}{1 - v}\left[\left(2\frac{\alpha_{n+1}}{\alpha_1} - v^n\right)^{-\frac{1}{n}} - 1\right] = \eta_{n+1} \tag{46}$$

Replacing v by the variable x and η_{n+1} by the variable y, and let $f(x, y)$ and $\eta(x)$ to be the expression on the left-hand side of (45) and (46) respectively; i.e.

$$f(x, y) = x^n + \left(1 - y + \frac{y}{x}\right)^{-n} - 2\frac{\alpha_{n+1}}{\alpha_1} \tag{47}$$

$$\eta(x) = \frac{x}{1 - x}\left[\left(2\frac{\alpha_{n+1}}{\alpha_1} - x^n\right)^{-\frac{1}{n}} - 1\right] \tag{48}$$

By construction, we have $f(x, \eta(x)) = 0$, for $x > 0$.

Let $\gamma = (\alpha_{n+1}/\alpha_1)^{1/n}$ and $a = [P_d(n)]^{1/n}$. Then $\eta(\gamma) = 1$ and $\eta(x) > 1$ for $x \in (\gamma, 1)$, $\eta(x) < 1$ for $x \in (0, \gamma)$.

In (i), note that $f(a, \eta(a)) = 0$ and $a \in (0, 1)$. Moreover f is strictly increasing as a function of x for any fixed $y > 0$. Hence $f(x, \eta(a)) < 0, \forall x \in (0, a)$. On the other hand, $f(x, y)$ is strictly decreasing as a function of y for any fixed $0 < x < 1$. This implies that $f(x, \eta) < 0, \forall \eta > \eta(a), x \in (0, a)$.

By definition, v can be written as

$$v^n = P_d(n)\sum_{i=0}^{n}\frac{A_d(n, i)}{P_d(n)}\frac{1}{1 + r(n, 0)\beta_n^i\Delta t}$$

Since all the short rates up to time period n are assumed to be positive, it follows from the convex property that $v \in (0, a)$. Hence there exist no positive interest rates in $(n + 1)$th period when the yield volatility $\hat{\sigma}_y(n + 1)$ satisfies that $\eta = \exp[2\hat{\sigma}_y(n + 1)\sqrt{\Delta t})] > \eta(a)$. Thus (i) is proved.

To prove part (ii), we consider the equation in terms of u and assume that $\eta < 1$ (by symmetry). Let $b = [P_u(n)]^{1/n} \in (0, 1)$. Then the proof is similar for

$$\eta = \exp[-2\hat{\sigma}_y(n + 1)\sqrt{\Delta t}] > \eta(b)$$

which is equivalent to

$$\hat{\sigma}_y(n + 1) < \frac{1}{2\Delta t}\log\left(\frac{1}{\eta(b)}\right)$$