Published in *Quantitative Finance* Vol. 15, No. 3 (2015): 443-454

A Generalized Procedure for Building Trees for the Short Rate and its Application to Determining Market Implied Volatility Functions

John Hull and Alan White

Joseph L. Rotman School of Management
University of Toronto
105 St George Street
Toronto M5S 3E6
Canada

Hull: 416 978 8615
hull@rotman.utoronto.ca
corresponding author

White: 416 978 3689
awhite@rotman.utoronto.ca

January 2014
This Version: June 2014

Abstract

One-factor no-arbitrage models of the short rate are important tools for valuing interest rate derivatives. Trees are often used to implement the models and fit them to the initial term structure. This paper generalizes existing tree building procedures so that a very wide range of interest rate models can be accommodated. It shows how a piecewise linear volatility function can be calibrated to market data and, using market data from days during the period 2004 to 2013, finds that the best fit to cap prices is provided by a function remarkably similar to that estimated by Deguillaume *et al* (2013) from historical data.

Key words: Interest Rate Models, Short Rate, Trees, Derivatives
JEL Classification: G13

A Generalized Procedure for Building Trees for the Short Rate and its Application to Determining Market Implied Volatility Functions

1. Introduction

One-factor models of the short rate, when fitted to the initial term structure, are widely used for valuing interest rate derivatives. Binomial and trinomial trees provide easy-to-use alternatives to finite difference methods for implementing these models.[1] Once the complete term structure has been calculated at each node, the tree can be used to value a wide range of derivatives or as a tool for simulating the evolution of the term structure. The latter is useful for some applications such as calculating the credit value adjustment for a portfolio.

Many authors show how trees can be built for particular short-rate models in such a way that they are consistent with the initial term structure of interest rates. Examples of binomial tree models are Ho and Lee (1986), Black *et al* (1990), Black and Karasinski (1991), and Kalotay *et al* (1993). Hull and White (1994, 1996) show how a trinomial tree can be constructed when the short rate, or some function of the short rate, is assumed to follow an Ornstein-Uhlenbeck process with a time-dependent reversion level.

The most popular short-rate models for derivative pricing are Ho and Lee (1986), Hull and White (1990), Black *et al* (1990), Black and Karasinski (1991), and Kalotay *et al* (1993). These models are used largely because it is relatively simple to build a tree for them. Indeed, in some cases the models are described by their authors entirely in terms of the tree that can be built. However, the models are all unrealistic in some respects. In Hull−White and Ho−Lee, the short rate exhibits normal behavior for all values of the short rate while in the other three models it exhibits lognormal behavior for all values of the rate.[2] Hull-White and Ho-Lee have the disadvantage that

---

[1] Complications in the use of finite difference methods are a) determining the boundary conditions and b) adapting the procedure so that the initial term structure is matched. However, these problems are not insurmountable since a trinomial tree can be regarded as a particular implementation of the explicit finite difference method.

[2] Ho and Lee (1986) is the particular case of Hull and White (1990) where there is no mean reversion of the short rate. Black et al (1990) is a particular case of Black and Karasinski (1991) where there is a relation between the drift rate and the volatility of the short rate. Kalotay *et al* (1993) can be viewed as a particular case of either Black and Karasinski (1991), or Black *et al* (1990) where the reversion rate of the logarithm of the short rate is zero.

they allow interest rates to become negative (and unfortunately the models give quite high probabilities of negative rates in the low-interest-rate environments experienced in many countries recently.) Rates are always positive in the other three models, but these models do not fit observed market prices well in low interest rate environments. (A change in the short rate from 10% to 20% has the same probability as a change from 20 basis points to 40 basis points.) A number of other tree-building procedures have recently been proposed for particular short rate models. For example, Hainaut and MacGilchrist (2010) show how to construct an interest rate tree when the short rate is driven by the normal inverse Gaussian process and Beliaeva and Nawalkha (2012) show how to construct trees for a constant elasticity of variance model when there are jumps.

The parameters of short rate models are typically chosen so that the prices of standard interest rate options (the "calibrating instruments") are matched as closely as possible. If the interest rate derivative being valued is similar to the calibrating instruments, the calculated price (and perhaps even the Greek letters) may not be sensitive to the model being used. But as the derivative becomes "more exotic" the model being used becomes progressively more important. Choosing the right model is also clearly important if the model is used (after a change from the Q- to the P-measure) to simulate the future evolution of the term structure for risk management purposes.

In many applications it is therefore clearly less than ideal to allow a particular technology for building interest rate trees to determine the short-rate model that is used. The short rate model should be chosen so that it is able to fit market prices and is consistent with empirical research on the historical behavior of rates. This paper provides a way this can be done. We show how trinomial tree procedures proposed by Hull and White (1994, 1996) can be extended so that they can be used for a much wider class of short rate models than those originally considered by the authors. Virtually any reasonable drift and volatility function can be accommodated. We first implement a simple model where enough information is presented to allow the reader to replicate the results. We then present a more elaborate model to illustrate the convergence characteristics of the procedure in the low interest rate environment at the end of 2013. Finally, we calibrate the model to market data on interest rate caps between 2004 and 2013. Interestingly, we find that market prices are consistent with recent empirical research by Deguillaume *et al* (2013).

2. Review of Hull–White Trinomial Trees

Hull and White (1994, 1996) consider models of the form

$$dx = [\theta(t) - ax]dt + \sigma\, dz \tag{1}$$

where x is some function $f(r)$ of the short rate r, a and σ are constants, and dz is a Wiener process.[3] Hull and White (1990) is the particular case of the model in equation (1) when $f(r) = r$ and Ho and Lee (1986) is the particular case of the model where $f(r) = r$ and a = 0. Black and Karasinski (1991) is the particular case of the model in equation (1) where $f(r) = \ln(r)$ and Kalotay et al (1993) is the particular case of the model where $f(r) = \ln(r)$ and a = 0. The parameter a is the reversion rate and $\theta(t)/a$ is a time-dependent reversion level with $\theta(t)$ chosen to fit the initial term structure. The Hull–White approach involves constructing a trinomial tree for x^* where

$$dx^* = -ax^*dt + \sigma dz$$

Suppose that the step size is Δt. The tree branching and probabilities on the tree are chosen so that the mean and variance of the change in x^* in time Δt are $-ax^*\Delta t$ and $\sigma^2\Delta t$, respectively. The vertical spacing between nodes, Δx^*, is set equal to $\sigma\sqrt{3\Delta t}$. The normal trinomial branching process in Figure 1a is used for all nodes that are less than $0.184/(a\Delta t)$ nodes from the center of the tree. The branching for the first node above $0.184/(a\Delta t)$ is Figure 1b and the branching for the first node below $-0.184/(a\Delta t)$ is Figure 1c.

A new variable $x = x^* + \phi(t)$ is then defined. The process followed by x is

$$dx = [\theta(t) - ax]dt + \sigma\, dz$$

where $\theta(t) = \phi'(t) + a\phi(t)$. This is the process in equation (1). We can therefore implement the model in equation (1) by searching step-by-step through the x^* tree to find the function $\phi(t)$ that correctly matches the initial term structure. In practice, we work forward from time zero shifting

[3] The model can be extended to allow a and σ to be functions of time. This allows volatilities of caps and swaptions to be matched more precisely, at the expense of the model becoming highly nonstationary. See, for example, Hull and White (2001).

all the nodes at time $i\Delta t$ by α_i so that a zero-coupon bond maturing at time $(i+1)\Delta t$ is correctly priced. The value of r at a node is calculated from the value of x at the node as $f^{-1}(x)$. Arrow Debreu prices for all nodes are calculated as the tree is constructed.

Arguments along the lines of those in Ames (1977) can be used to prove convergence for this type of tree. It is important to note that the rate on the tree is the Δt-period rate, expressed with continuous compounding. It is not the instantaneous short rate.[4] This distinction proves to be important when models such as Ho and Lee (1986) and Hull and White (1990) are used in conjunction with analytic expressions for bond prices and European options.

For $f(r)$ other than r, the drift of r can be a strange function r. Ito's lemma gives the process for r as

$$dr = \left\{ \left[\theta(t) - ax \right] h(x) + h'(x)\sigma^2/2 \right\} dt + \sigma\, h(x)\, dz$$

where $h(x) = dr/dx$. The shape of the volatility function, $h(x)$, determines the drift of r. If the volatility function has a discontinuous first derivative the drift of r is discontinuous. In the case of the Black-Karasinski model the process for r is

$$dr = r\left\{ \theta(t) - a\ln(r) + \sigma^2/2 \right\} dt + \sigma r\, dz$$

The drift of r for a particular value of θ is shown in Figure 2.

It is tempting to modify the drift in equation (1) so that the drift of r has better properties. However, this is not possible. The Hull and White (1994, 1996) tree-building procedure works only when the drift of x is linear in x. The only justification for accepting the unusual drift for r is that a tree for x (and therefore for r) can be constructed fairly easily. In the next section we propose an alternative procedure for building short-rate trees that allows us to select virtually any one-factor model for the short rate.

[4] Strictly speaking, the numeraire is not the money market account except in the limit as Δt tends to zero. Within each time step the numeraire is the price of a zero coupon bond maturing at the end of the time step.

3. A Generalization

Assume that the short rate r follows the process

$$dr = \left[\theta(t) + F(r)\right]dt + G(r)dz \tag{2}$$

where $\theta(t)$ is a function of time chosen so that the model fits the initial term structure while $F(r)$ and $G(r)$ are functions determining the drift and volatility of r.[5] We require $G(r)$ to be continuously differentiable. Define

$$x = f(r) = \int \frac{dr}{G(r)}$$

so that

$$\frac{dx}{dr} = \frac{1}{G(r)}$$

The process followed by x is

$$dx = H\left(x,t\right)dt + dz$$

where the drift of x, $H\left(x,t\right)$, is

$$H\left(x,t\right) = \frac{\theta\left(t\right) + F\left(r\right)}{G\left(r\right)} - \frac{1}{2}G'\left(r\right) \quad \text{and} \quad r = f^{-1}\left(x\right) \tag{3}$$

We build a tree for x (or equivalently a tree for r) by considering points on a grid that are equally spaced in x and time. We denote the x-spacing by Δx and the time spacing by Δt.[6] The main differences between the generalized procedure we describe here and the procedure in Section 2 are

---

[5] The methodology can be extended to more general models of the form

$$dr = F\left(r,t\right)dt + G\left(r,t\right)dz$$

where the dependence of the F function on time involves a function $\theta(t)$ that is chosen to fit the term structure.

[6] For ease of exposition we assume constant time steps. In practice, the time step may not be constant because, when valuing a derivative, we usually wish to have nodes on each payment date. The procedure we describe can be adjusted as in Hull and White (2001) to accommodate varying time steps.

1. In the generalized procedure, we consider values for x that lie on a fixed rectangular grid throughout the numerical procedure. We do not shift points on the grid to match the term structure. (As explained at the end of Section 2, the latter approach only works for particular short rate models.)
2. The branching process from a node at time $i\Delta t$ may be different from any of the branching processes in Figure 1. We always branch to a triplet of three adjacent nodes at time $(i+1)\Delta t$, but the middle node can correspond to any of the values of x considered by the grid.
3. As in the case of the procedure in Section 2, we move forward in time through the tree searching iteratively for values of $\theta(t)$ that match the initial term structure. However, in the case of the generalized procedure, the branching processes and the branch probabilities at time t depend on the value of $\theta(t)$.

We now provide details. The x-values considered by the tree are

$$\dots x_0 - 2\Delta x,\ x_0 - \Delta x,\ x_0,\ x_0 + \Delta x,\ x_0 + 2\Delta x \dots$$

where $\Delta x = \sqrt{3\Delta t}$, $x_0 = f(r_0)$, and r_0 is the initial Δt interest rate.

The node at time $i\Delta t$ for which $x = x_0 + j\Delta x$ will be denoted by the (i, j) node. The initial x-node is therefore (0, 0). We define $j_d(i)$ and $j_u(i)$ and the lowest and highest values of j that can be reached at time $i\Delta t$.

For a particular value of θ at time $i\Delta t$, the branching process from node (i, j) is determined as follows. Define the first and second moment of x at time $(i+1)\Delta t$ conditional on x being at node (i, j) at time $i\Delta t$ as m_1 and m_2, respectively. We set

$$m_1 = f\big(r_j + \big(\theta + F(r_j) - 0.5G(r_j)G'(r_j)\big)\Delta t\big) \quad \text{and} \quad m_2 = \Delta t + m_1^2$$

where r_j is the Δt rate at node (i, j). By expanding $f(r)$ in the expression for m_1 in a Taylor series and noting that $f'(r) = 1/G(r)$ we see from equation (3) that, in the limit as Δt tends to zero, it gives the correct first moment for x. The variance of x is in the limit Δt. The expression for the second moment is therefore also correct in the limit as Δt tends to zero.

The nodes that can be reached from node (i, j) are $(i+1, j^*-1)$, $(i+1, j^*)$, and $(i+1, j^*+1)$ where j^* is chosen so that $x_0 + j^*\Delta x$ is as close as possible to m_1. This means that

$$j^* = \mathrm{int}\left(\frac{m_1 - x_0}{\Delta x} + 0.5\right)$$

Define p_d, p_m, and p_u as the probabilities of branching from node (i, j) to nodes $(i+1, j^*-1)$, $(i+1, j^*)$, and $(i+1, j^*+1)$, respectively. We choose the probabilities so that the moments are matched. This means that

$$\begin{aligned}
p_u &= \frac{\Delta t + \left(m_1 - x_0 - j^*\Delta x\right)^2}{2\Delta x^2} + \frac{m_1 - x_0 - j^*\Delta x}{2\Delta x} \\
p_d &= \frac{\Delta t + \left(m_1 - x_0 - j^*\Delta x\right)^2}{2\Delta x^2} - \frac{m_1 - x_0 - j^*\Delta x}{2\Delta x} \\
p_m &= 1 - p_u - p_d
\end{aligned}$$

Because $\Delta x = \sqrt{3\Delta t}$ and $m_1 - x_0 - j^*\Delta x \le 0.5\Delta x$, these probabilities are always positive.

Define $Q_{i,j}$ as the Arrow Debreu price of node (i, j). This is the value of a derivative that pays off one if node (i, j) is reached and zero otherwise. We start at time zero, setting $Q_{0,0} = 1$ and $j_d(0) = j_u(0) = 0$. We then successively consider the nodes at times $\Delta t,\ 2\Delta t,\ 3\Delta t, \ldots$

At time $i\Delta t$ we choose a trial value of $\theta\left(\left(i-1\right)\Delta t\right)$. This is used to determine branching probabilities for each node at time $(i-1)\Delta t$ as well as $j_d(i)$ and $j_u(i)$. The $Q_{i,j}$ are determined for all j using

$$Q_{i,j} = \sum_{k=j_d(i-1)}^{j_u(i-1)} Q_{i-1,k}\, q(k, j) \exp(-r_k \Delta t)$$

where $q(k, j)$ is the probability of moving from node $(i-1, k)$ to (i, j) and $r_k = f^{-1}(x_0 + k\Delta x)$.[7] The price of a zero-coupon bond with maturity $(i+1)\Delta t$ is then determined as

$$P_{i+1} = \sum_{j=j_d(i)}^{j_u(i)} Q_{i,j} \exp\left(-r_j \Delta t\right)$$

We iterate searching for the value of $\theta\left(\left(i-1\right)\Delta t\right)$ for which this bond price is consistent with the initial term structure of interest rates. Once the correct value of $\theta\left(\left(i-1\right)\Delta t\right)$ is found, $j_d(i)$ and $j_u(i)$, the final branching probabilities, and the $Q_{i,j}$ are stored. We are then ready to move on to consider the nodes at time $(i+1)\Delta t$.

Continuing in this way the complete tree is constructed. It is used to value interest rate derivatives in an analogous way to other interest rate trees. The complete term structure can be calculated at each node by rolling back through the tree to calculate zero-coupon bonds with different maturities and recording their values at each node.

Occasionally when $\theta\left(\left(i-1\right)\Delta t\right)$ is being determined, the branching process oscillates during the iterative procedure in such a way that the $(i+1)\Delta t$ bond price cannot be exactly matched.[8] We handle this problem by freezing the current branching process between times $(i-1)\Delta t$ and $i\Delta t$ when the non-convergence is observed and then repeating the iterative procedure to determine $\theta\left(\left(i-1\right)\Delta t\right)$. This does not lead to negative branching probabilities because p_u, p_m, and p_d remain positive for a wide range of positions of the central node relative to mean value of x. Specifically, they are positive when the distance between the mean value of x and the central node $(= m_1 - x_0 - j^*\Delta x)$ is less that $0.758\Delta x$.

Once the interest rate derivative has been valued it is usually necessary to calculate Greek letters. This should be done by freezing the final branching process for the whole tree and making small changes to the relevant variables. This procedure minimizes the impact of noise on the results.

---

[7] For each k only three values of $q(k, j)$ are non-zero.

[8] This ceases to happen altogether as Δt tends to zero.

One issue usually needs to be addressed for the model in equation (2). If the drift of r is negative for some values of t (as can happen, for example, when forward rates decline), the short rate may become negative. If x or $G(r)$ is undefined for zero or negative interest rates this causes problems. We handle this by making a small adjustment to the drift of r so that it never falls below zero. We do this by setting

$$m_1 = f\left[\max\left(r_j + \left(\theta + F(r_j) - 0.5G(r_j)G'(r_j)\right)\Delta t, \varepsilon\right)\right]$$

We find that $\varepsilon = 0.0001$ works well.[9]

Given the trend toward OIS discounting, it is often necessary to consider the behavior of two rates simultaneously when valuing interest rate derivatives. Hull and White (2015) propose a way in which a procedure for constructing a tree for the rate used for discounting, such as the one given here, can be augmented with another procedure that calculates at each node the expected value of the rate used to determine payoffs.

Example 1

For a detailed illustration of the procedure, we consider a simple example. We choose $F(r) = -ar$ and $G(r) = \sigma r$ with $a > 0$ and $\sigma > 0$ so that the process for the short rate is

$$dr = \left[\theta(t) - ar\right]dt + \sigma r dz$$

For this example, $x = \ln(r)/\sigma$ and $r = e^{\sigma x}$. Figure 3 shows a two-year tree for the model when the initial term structure is given in Table 1, Δt = 0.5 years, a = 0.2 and σ = 0.15. The values of $\theta(t)$ together with the $j_d(i)$ and $j_u(i)$ are shown in Table 2. Table 3 shows the branching probabilities and Table 4 shows the Arrow-Debreu prices and the Δt-period rate at each node.

Example 2

As a second more realistic illustration of the tree-building procedure, we let the $G(r)$ function be

---

[9] For some currencies it may be desirable to let the interest rate become slightly negative. Suppose that it is considered that the rate could become $-e$ The simplest way of handling this is to assume that the procedures proposed here be used for $r+e$ instead of r.

$$
G(r)=\begin{cases} s\left[2r/r_1-(r/r_1)^2\right] & \text{when} \quad r\le r_1 \\ s+K(r-r_1)^2 & \text{when} \quad r_1<r\le r_2 \\ \alpha+\beta r & \text{when} \quad r>r_2 \end{cases} \tag{4}
$$

where K is a positive constant. This function is always continuous and differentiable at $r = r_1$. To ensure continuity and differentiability at $r = r_2$, we set $K=\beta/\left[2(r_2-r_1)\right]$ and $\alpha=s+K(r_2-r_1)^2-\beta r_2$. The function was chosen because it is similar to the function derived for historical short-rate movements by Deguillaume *et al* (2013). (The latter is discussed further in the next section.) The drift was as in Example 1 with $a = 0.05$. We set $r_1 = 0.02$, $r_2 = 0.1$, $s = 0.02$, and $\beta = 0.2$ so that $K = 1.25$ and $\alpha = 0.008$. The volatility function is shown in Figure 4. The term structure was that on December 2, 2013 and is shown in Table 5.

From equation (4), x is defined as follows

$$
x=\begin{cases} \dfrac{r_1}{2s}\ln\dfrac{r}{2r_1-r} & \text{when} \quad r\le r_1 \\ \dfrac{1}{\sqrt{sK}}\tan^{-1}\left((r-r_1)\sqrt{K/s}\right) & \text{when} \quad r_1<r\le r_2 \\ \dfrac{1}{\beta}\ln(\alpha+\beta r)+C & \text{when} \quad r>r_2 \end{cases}
$$

where

$$
C=-\frac{1}{\beta}\ln(\alpha+\beta r_2)+\frac{1}{\sqrt{sK}}\tan^{-1}\left((r_2-r_1)\sqrt{K/s}\right)
$$

The constants of integration are chosen so that x is continuous.

The inverse function is

$$r = \begin{cases} \dfrac{2r_1}{1+\exp\left(-2sx/r_1\right)} & \text{when} \quad x \le 0 \\ r_1 + \sqrt{s/K}\,\tan\left(x\sqrt{sK}\right) & \text{when} \quad 0 < x \le x_1 \\ \dfrac{\exp\left(\beta(x-C)\right)-\alpha}{\beta} & \text{when} \quad x > x_1 \end{cases}$$

where

$$x_1 = \frac{1}{\sqrt{sK}} \tan^{-1}\left(\left(r_2 - r_1\right)\sqrt{K/s}\right)$$

Table 6 shows how the values calculated for annual-pay caps converge as the number of time steps is increased.

4. The Volatility Function Implied By Market Data

The attraction of the generalized model is that users are able to specify the drift and the volatility function, $G(r)$, in any way they like.[10] In particular, the prices of actively traded options can be used to determine the volatility function. If $G(r)$ is defined by some set of parameters, the model can be calibrated to market data by choosing the parameter values so that the model fits observed prices as well as possible. For example, a mean reverting drift could be specified and the volatility function could be set to $G(r)=\sigma\sqrt{r}$. In this case, the calibration would find the value of σ that best matches observed option prices.

We choose to specify the volatility in a more general way. We let $G(r)$ be a piecewise linear function of r defined by a set of specified corner points, r_i, for $i = 1, \ldots, n$ and the values of the function at each corner point, s_i. Since this function is not continuously differentiable, it is necessary to develop a procedure for "rounding" the corner points. The details of this procedure appear in the Appendix. To avoid negative interest rates $G(0)$ is set to zero. The model is then calibrated to market data by choosing the values of s_i that result in a best fit to observed market prices. The market data therefore defines the shape of the $G(r)$ function.

To illustrate this process we calibrate the model

$$dr=\left[\theta(t)-ar\right]dt+G(r)\,dz \tag{5}$$

with $a = 0.05$ to the prices of 10-year caps with different strikes observed on December 2, 2013. The term structure is that given in Table 5 and the cap volatilities and prices are given in Table 7. We chose $n = 7$ and set the values of r_i equal to 1%, 2%, 3%, 4%, 5%, 6%, and 10% . The "goodness-of-fit" objective function we used was

$$\sum_{i=1}^{N}\frac{\left(U_i-V_i\right)^2}{U_i}$$

[10] The ability to define the drift of the process is particularly important if the model is to be adapted to model the evolution of real (as opposed to risk-neutral) interest rates for risk management purposes. (See Hull, Sokol, and White (2014).

where N is the number of different caps used in the calibration (as indicated in Table 7, N = 10 for this test), and U_i and V_i are the market price and model price of the ith cap. An optimizer was then used to find the s_i that minimized the objective function.

The fitted cap prices and the pricing errors are shown in Table 7 and the best fit set of s_i that result are shown in Table 8[11]. The resulting $G(r)$ is shown in Figure 5. Because of the large number of degrees of freedom in the calibration, the fit to the market prices is very good. Unfortunately, the type of high dimensional optimization involved in determining the best fit $G(r)$ function when n = 7 is both time consuming and difficult. However, an examination of the $G(r)$ function illustrated in Figure 5 shows that a simpler functional form may be acceptable. To explore this possibility we recalibrated the model using a $G(r)$ with only three corner points: 1%, 5% and 10%. The fitted cap prices and the pricing errors are shown in Table 9 and the best fit set of s_i that result is shown in Table 10. The resulting $G(r)$ is shown in Figure 5.

The best fit $G(r)$ for the December 2, 2013 10-year caps is similar to the empirical observations of Deguillaume *et al* (2013). This paper shows that the short rate exhibits approximate lognormal behavior when the rate is low or high. For intermediate values of the rate, the behavior is approximately normal. The process derived from historical data by Deguillaume *et al* (2013) is

$$dr = \ldots + \sigma(r)\,dz$$

where

$$\sigma(r) = \begin{cases} s\,r/r_L & \text{when} \quad r \le r_L \\ s & \text{when} \quad r_L < r \le r_U \\ s + \beta(r - r_U) & \text{when} \quad r > r_U \end{cases}$$

This function is illustrated in Figure 6.

The Deguillaume *et al* results are remarkably consistent across currencies and time periods. Because they are based on historical data, they apply to rate changes in the real world. However,

[11] Table 7 also shows the pricing errors that arise if the Hull-White (normal) or Black-Karasinski (lognormal) models are calibrated to the data. These models both have a systematic bias in their fit to the data. The Hull-White model tends to over-price low strike options and under-price high strike options. The reverse is true for the Black-Karasinski model.

Girsanov's well-known theorem shows that volatilities are unaffected by a change of measure. We should therefore expect the volatility of the short rate as a function of the short rate to be the same in both the real world and the risk-neutral world. Our results for the December 2013 10-year caps suggest that the volatility function has the same shape in the two worlds

The Deguillaume *et al* results are based on the long run average behavior of interest rates. To explore the longer-run average behavior of implied $G(r)$ functions we calibrated the model in equation (5) to cap prices observed between 2004 and 2013. Market data for caps observed on the last trading day of March in each year were used. To minimize the effect on the calibration of very high and very low strike options the cap quotes used for a particular maturity were those where the cap rate was within 1.5 standard deviations of the at-the-money strike price.[12] Because of this filtering the short-dated caps that are included in the calibration generally have fairly low strike prices. We were concerned that this might affect to calibration process, particularly the portion of the $G(r)$ function that applies to high rates. As a result we divided the cap data into short-term caps, caps with maturities between 1 and 10 years, and long-term caps, caps with maturities between 10 and 30 years. The $G(r)$ function for each year was estimated for these two groups separately. In each calibration, the model was fitted to about 50 caps with different strikes and maturities.

The choice of the corner points in the parameterization of $G(r)$ was determined by experimentation. We first tried the corner points used when calibrating to the December 2013 data, 1% and 5%. We then tried 2% and 6% and ultimately decided to use 1.5% and 6% on the grounds that the goodness of fit for these corner points was slightly better on average than for the alternatives. The best fit parameters for each of the calibrations, as well as the average parameter values averaged over the ten years of observations, are shown in Table 11 and illustrated in Figures 7 and 8. In every year the $G(r)$ function has the same shape as the Deguillaume *et al* function. Further, there is a high degree of similarity between the results for calibration to short-term caps and calibration to long-term caps in each year. Most of the deviation is observed in the value of $G(r)$ for high values of r where the short term caps provide little information.

[12] The standard deviation was the average cap volatility for the maturity being considered multiplied by the square-root of the maturity.

The volatility structure implied by cap prices is similar to that observed in the real world. The structures calibrated from cap prices provide a measure of the volatility structure at a point in time. As the results show, this changes from year to year and probably from day to day. Since we cannot observe the real world volatility structure over short periods of time the best we can say is that these results are consistent with Girsanov's theorem. Overall, our results are supportive of the conjecture that the volatility function used (implicitly or explicitly) by market participants when pricing interest rate caps is similar to the volatility function derived by Deguillaume *et al*. What is more, this was true even before the Deguillaume *et al* research was first available as a working paper.

5. Conclusions

Interest rate trees are useful tools for implementing models of the short rate. Researchers have tended to use short rate models that have a particular form because there are well established procedures in the literature for constructing trees for these models. This paper greatly expands the range of models that can be used. This means that the choice of an interest rate model can be related more closely to the behavior of interest rates. We have shown how our tree-building technology can be used for a general model where the volatility function is approximately piecewise linear and the drift is mean-reverting. When we calibrate the model to market data, we find that a volatility function close to that estimated by Deguillaume *et al* (2013) from historical data gives the best fit. Girsanov's theorem shows that volatilities are invariant to a change of measure. The paper therefore provides important support for the Deguillaume et al research findings. It also is suggestive of the volatility function that should be used to value non-standard interest rate options.

References

Ames, William F. (1977), *Numerical Methods for Partial Differential Equations*, New York: Academic Press.

Beliaeva, Natalia, Sanjay K. Nawalkha (2012), "Pricing American Interest Rate Options Under the Jump Extended Constant-Elasticity-of-Variance Short Rate Models," *Journal of Banking and Finance*, Vol. 36, No.1, 151-163.

Black, Fischer, Emanuel Derman, and William Toy (1990), "A One-Factor Model of Interest Rates and its Application to Treasury Bond Prices," *Financial Analysts Journal*, (January/February), 46, pp. 33-39.

Black, Fischer and Piotr Karasinski (1991), "Bond and Option Pricing When Short Rates are Lognormal," *Financial Analysts Journal* (July/August), 47, pp. 52-59.

DeGuillaume, Nick, Riccardo Rebonato, and Andrei Pogudin (2013), "The Nature of the Dependence of the Magnitude of Rate Moves on the Level of Rates: A Universal Relationship," *Quantitative Finance*, 13, 3, pp. 351-367.

Hainaut, Donatien and Renaud MacGilchrist, "An Interest rate Tree Driven by a Lévy Process," *Journal of Derivatives*, Vol. 18, No. 2 (Winter), 33-45.

Ho, Thomas S.Y. and Sang-B. Lee (1986), "Term Structure Movements and Pricing Interest Rate Contingent Claims," *Journal of Finance*, 41 (December), pp. 1011-1029

Hull, John and Alan White (1990), "Pricing interest-rate derivative securities", *The Review of Financial Studies*, Vol 3, No. 4 pp. 573-592

Hull, John and Alan White (1994), "Numerical procedures for implementing term structure models I," *Journal of Derivatives*, 2 (Fall), pp 7-16

Hull, John and Alan White (1996), "Using Hull-White interest rate trees," *Journal of Derivatives*, Vol. 3, No. 3 (Spring), pp 26-36.

Hull, John and Alan White (2001), "The General Hull-White Model and SuperCalibration," *Financial Analysts Journal*, 57, 6, (Nov-Dec) 2001., pp34-43

Hull, John and Alan White (2015), “OIS Discounting, Interest Rate Derivatives, and the Modeling of Stochastic Interest Rate Spreads,” *Journal of Investment Management*. 13, 1, pp 1-20.

Hull, John, Alexander Sokol, and Alan White (2014), “Joint Measure Models of the Short Rate” *Risk*, 27, 10 (October).

Kalotay, Andrew J., George O. Williams, and Frank J. Fabozzi (1993), “A Model for Valuing Bonds and Embedded Options,” *Financial Analysts Journal*, 49, 3 (May-June), pp. 35-46.

Appendix

In this appendix we describe how to fit an arbitrary volatility function, $G(r)$. Suppose that $G(0) = 0$, $G(r_1) = s_1$, $G(r_2)= s_2$, $G(r_3) = s_3$ and so on. If we fit a piecewise linear function to this data then

$$
\begin{gathered}
G(r) = a_i + b_i r \quad \text{for} \quad r_i \le r < r_{i+1} \\
a_i = \frac{s_i r_{i+1} - s_{i+1} r_i}{r_{i+1} - r_i} \quad \text{and} \quad b_i = \frac{s_{i+1} - s_i}{r_{i+1} - r_i}
\end{gathered}
\tag{A1}
$$

In order for the function to be continuous we require that the values on both sides of the corner points be the same. As a result

$$
a_{i+1} = a_i + (b_i - b_{i+1}) r_{i+1} \tag{A2}
$$

The first derivative of the function is

$$
G'(r) = b_i \quad \text{for} \quad r_i \le r < r_{i+1}
$$

This derivative is discontinuous at the corner points.

In order to make the piecewise linear function be continuous and have a continuous derivative we insert a quadratic, $x_{i+1} + y_{i+1} r + z_{i+1} r^2$, between $r_{i+1} - \varepsilon$ and $r_{i+1} + \varepsilon$. If $G(r)$ is to be continuous

$$
\begin{aligned}
a_i + b_i (r_{i+1} - \varepsilon) &= x_{i+1} + y_{i+1}(r_{i+1} - \varepsilon) + z_{i+1}(r_{i+1} - \varepsilon)^2 \\
a_{i+1} + b_{i+1}(r_{i+1} + \varepsilon) &= x_{i+1} + y_{i+1}(r_{i+1} + \varepsilon) + z_{i+1}(r_{i+1} + \varepsilon)^2
\end{aligned}
$$

If $G(r)$ is to have a continuous first derivative

$$
\begin{aligned}
b_i &= y_{i+1} + 2 z_{i+1}(r_{i+1} - \varepsilon) \\
b_{i+1} &= y_{i+1} + 2 z_{i+1}(r_{i+1} + \varepsilon)
\end{aligned}
$$

Making use of equation (A2) these conditions can be solved for z, y and x

$$
\begin{aligned}
z_{i+1} &= \frac{b_{i+1} - b_i}{4\varepsilon} \\
y_{i+1} &= b_i - 2z_{i+1}\left(r_{i+1} - \varepsilon\right) \\
x_{i+1} &= a_i + z_{i+1}\left(r_{i+1} - \varepsilon\right)^2
\end{aligned}
\tag{A3}
$$

The resulting continuously differentiable function is:

$$
G(r) = \begin{cases} a_0 + b_0 r & 0 \le r \le r_i - \varepsilon \\ x_1 + y_1 r + z_1 r^2 & r_1 - \varepsilon \le r \le r_1 + \varepsilon \\ a_1 + b_1 r & r_1 + \varepsilon \le r \le r_2 - \varepsilon \\ x_2 + y_2 r + z_2 r^2 & r_2 - \varepsilon \le r \le r_2 + \varepsilon \\ \cdots & \cdots \end{cases}
$$

where the coefficients are given in equations (A1), (A2) and (A3).

In the linear regions the $x(r)$ function is

$$
x(r) = \begin{cases} \ln\left(a + br\right)/b + C & b \ne 0 \\ r/a + C & b = 0 \end{cases}
$$

In the quadratic regions the $x(r)$ function is

$$
x(r) = \begin{cases} \dfrac{2}{\sqrt{D}} \tan^{-1}\left(\dfrac{2zr + y}{\sqrt{D}}\right) + C & D > 0 \\ \dfrac{-2}{2zr + y} + C & D = 0 \\ \dfrac{1}{\sqrt{D}} \ln\left(\left|\dfrac{2zr + y - \sqrt{D}}{2zr + y + \sqrt{D}}\right|\right) + C & D < 0 \end{cases}
$$

where $D = 4xz - y^2$. The constants of integration, C, are chosen to make $x(r)$ continuous.

In the linear regions the $r(x)$ function is

$$
r(x) = \begin{cases} \exp\left(b\left(x - C\right) - a\right)/b & b \ne 0 \\ a\left(x - C\right) & b = 0 \end{cases}
$$

In the quadratic regions the $r(x)$ function is

$$r(x)=\begin{cases} \dfrac{\tan\left(\dfrac{\sqrt{D}(x-C)}{2}\right)\sqrt{D}-y}{2z} & D>0 \\ \dfrac{-2/(x-C)+y}{2z} & D=0 \\ \left.\begin{matrix} \dfrac{\sqrt{D}(1+q)/(1-q)-y}{2z} \\ \text{or} \\ \dfrac{\sqrt{D}(1-q)/(1+q)-y}{2z} \end{matrix}\right\} & D<0 \end{cases}$$

where $q=\exp\left(\sqrt{D}(x-C)\right)$. When the discriminant, D, is negative only one of the two roots lies within the current region.

This procedure allows us to replicate almost any $G(r)$ function with high fidelity. The only condition that must be satisfied is that $r_{i+1}-r_i>2\varepsilon$.

Table 1

Zero Coupon Interest Rates for Example 1

Maturity (years)	Rate per annum (% cont. comp.)
0.5	5.0
1.0	6.0
1.5	7.0
2.0	7.5
3.0	8.5

Table 2

Values of Theta for Example 1

Time (years)	Minimum Node	Maximum Node	Theta
0.0	0	0	0.04980
0.5	1	3	0.05387
1.0	2	*5*	0.01817
1.5	1	6	0.03812
2.0	1	7	

Table 3

Probabilities of Transitions between Nodes on the Tree for Example 1

Time (years)	Start Node	End Node 1	2	3	4	5	6	7
0.0	0	0.2853	0.6275	0.0873				
0.5	1		0.4432	0.5097	0.0471			
0.5	2		0.0637	0.5826	0.3537			
0.5	3			0.1597	0.6665	0.1739		
1.0	2	0.1182	0.6549	0.2269				
1.0	3		0.1692	0.6666	0.1641			
1.0	4			0.2227	0.6563	0.1210		
1.0	5				0.2752	0.6330	0.0918	
1.5	1	0.1452	0.6646	0.1902				
1.5	2		0.2864	0.6268	0.0868			
1.5	3			0.4575	0.4970	0.0455		
1.5	4			0.0462	0.5028	0.4510		
1.5	5				0.0735	0.6054	0.3211	
1.5	6					0.1165	0.6539	0.2296

Table 4

Arrow Debreu Prices and the Interest Rate at each node for Example 1

Time (years)	0	1	2	3	4	5	6
0.0	0.2782	0.6120	0.0851				
0.5		0.1573	0.4945	0.2758	0.0142		
1.0	0.0179	0.1795	0.4084	0.2532	0.0401	0.0012	
1.5	0.0025	0.0612	0.3018	0.3330	0.1491	0.0128	0.0003
Node Rate	6.008%	7.220%	8.676%	10.426%	12.528%	15.055%	18.091%

Table 5

Zero coupon interest rates for Example 2 and cap calibration

Time (yrs)	Zero Rate
0.25	0.20%
0.5	0.22%
1	0.27%
2	0.38%
3	0.67%
4	1.08%
5	1.52%
6	1.93%
7	2.27%
8	2.55%
9	2.78%
10	2.98%
15	3.60%
20	3.85%
25	3.96%
30	4.02%

Table 6

Values of caps with annual payments on December 2, 2013 when volatility function is as shown in Figure 4, cap rate is 4%, and the principal is 100.

Steps	Cap Life (yrs)		
per year	10	20	30
1	8.56	21.49	29.11
2	7.84	20.18	27.38
5	7.65	20.00	27.26
10	7.67	20.11	27.43
15	7.63	20.03	27.33
20	7.64	20.01	27.29

Table 7

Calibrating a piecewise linear $G(r)$ function to the market prices of 10-year caps on December 2, 2013. The columns headed HW Price Difference and BK Price Difference show the errors that arise when the Hull-White (normal) or Black-Karasinki (lognormal) models are used. The term structure on that date is shown in Table 5. The best fit $G(r)$ function is shown in Table 8 and illustrated in Figure 5.

Cap Strike	Market Volatility	Market Price	Model Price	Price Difference	HW Price Difference	BK Price Difference
1.00%	50.75%	19.18%	18.87%	–0.31%	2.19%	–1.03%
2.00%	38.73%	14.06%	13.86%	–0.20%	1.89%	–1.27%
3.00%	32.30%	9.96%	9.99%	0.03%	1.60%	–0.89%
4.00%	30.15%	7.21%	7.12%	–0.09%	0.89%	–0.67%
5.00%	28.50%	5.19%	5.10%	–0.09%	0.28%	–0.39%
6.00%	26.50%	3.56%	3.68%	0.12%	–0.01%	0.05%
7.00%	25.72%	2.56%	2.63%	0.07%	–0.35%	0.17%
8.00%	25.50%	1.92%	2.01%	0.09%	–0.61%	0.19%
9.00%	25.55%	1.50%	1.46%	–0.04%	–0.76%	0.13%
10.00%	25.70%	1.20%	1.12%	–0.08%	–0.80%	0.10%

Table 8

The corner points of the best fit piecewise linear $G(r)$ function calibrated to the market data in Table 7. This function is illustrated in Figure 5.

r_i	0.0%	1.0%	2.0%	3.0%	4.0%	5.0%	6.0%	10.0%
s_i	0.00%	1.48%	1.68%	1.68%	1.80%	1.97%	2.33%	3.43%

Table 9

Results from calibrating a piecewise linear $G(r)$ function to the market prices of 10-year caps on December 2, 2013. The term structure on that date is shown in Table 5. The best fit $G(r)$ function is shown in Table 10 and illustrated in Figure 5. The market price is expressed as a percent of notional principal.

Cap Strike	Market Volatility	Market Price	Model Price	Price Difference
1.00%	50.75%	19.18%	19.01%	–0.17%
2.00%	38.73%	14.06%	14.07%	0.01%
3.00%	32.30%	9.96%	10.18%	0.22%
4.00%	30.15%	7.21%	7.23%	0.01%
5.00%	28.50%	5.19%	5.08%	–0.11%
6.00%	26.50%	3.56%	3.59%	0.04%
7.00%	25.72%	2.56%	2.58%	0.02%
8.00%	25.50%	1.92%	1.95%	0.03%
9.00%	25.55%	1.50%	1.41%	–0.09%
10.00%	25.70%	1.20%	1.09%	–0.11%

Table 10

The value of the best fit piecewise linear $G(r)$ function, s_i, calibrated to the market data in Table 9 at each of the corner points. This function is illustrated in Figure 5.

r_i	0.0%	1.0%	5.0%	10.0%
s_i	0.00%	1.62%	1.83%	3.48%

Table 11

The value of the best fit piecewise linear $G(r)$ function, s_i, calibrated to the market data in Table 9 at each of the corner points. The upper panel shows the s_i for the case in which the calibration set is short-term caps (illustrated in Figure 7). The middle panel shows the s_i for the case in which the calibration set is long-term caps (illustrated in Figure 8). The lower panel shows the difference between the two calibrations.

	Short-term Caps: 1- to 10-year maturity										
	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	Avg.
0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1.50%	1.59%	0.93%	1.05%	0.92%	1.30%	1.28%	1.61%	1.55%	1.87%	1.96%	1.41%
6.00%	1.91%	1.21%	1.05%	0.94%	1.35%	1.63%	1.70%	2.00%	2.10%	2.16%	1.60%
10.00%	2.50%	2.11%	2.07%	1.88%	1.57%	4.40%	2.55%	3.38%	3.37%	4.71%	2.85%
	Long-term Caps: 10- to 30-year maturity										
	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	Avg.
0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1.50%	1.49%	1.19%	1.16%	1.01%	1.28%	1.37%	1.48%	1.56%	1.84%	1.88%	1.43%
6.00%	1.59%	1.21%	1.16%	1.01%	1.30%	1.52%	1.48%	1.76%	1.89%	1.90%	1.48%
10.00%	1.95%	1.95%	1.80%	1.63%	1.68%	4.30%	2.54%	2.82%	3.36%	4.88%	2.69%
	Difference in Calibration Results: Long-term parameter less Short-term parameter										
	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	Avg.
0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1.50%	–0.10%	0.26%	0.11%	0.09%	–0.02%	0.09%	–0.13%	0.01%	–0.03%	–0.08%	0.02%
6.00%	–0.32%	0.00%	0.11%	0.07%	–0.05%	–0.11%	–0.22%	–0.24%	–0.21%	–0.26%	–0.12%
10.00%	–0.54%	–0.15%	–0.27%	–0.25%	0.12%	–0.09%	–0.01%	–0.57%	–0.02%	0.17%	–0.16%

Figure 1

Branching processes used in Hull and White (1994, 1996)

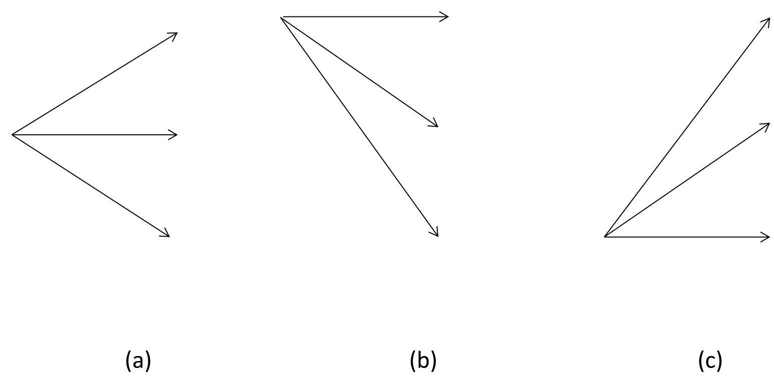

Figure 2

The drift of the short rate, *r*, as a function of *r* in Black and Karasinski (1991) when the volatility of *r* is 20%, the reversion rate, *a*, is 5% and the reversion level, $\theta(t)/a$, is ln(0.03).

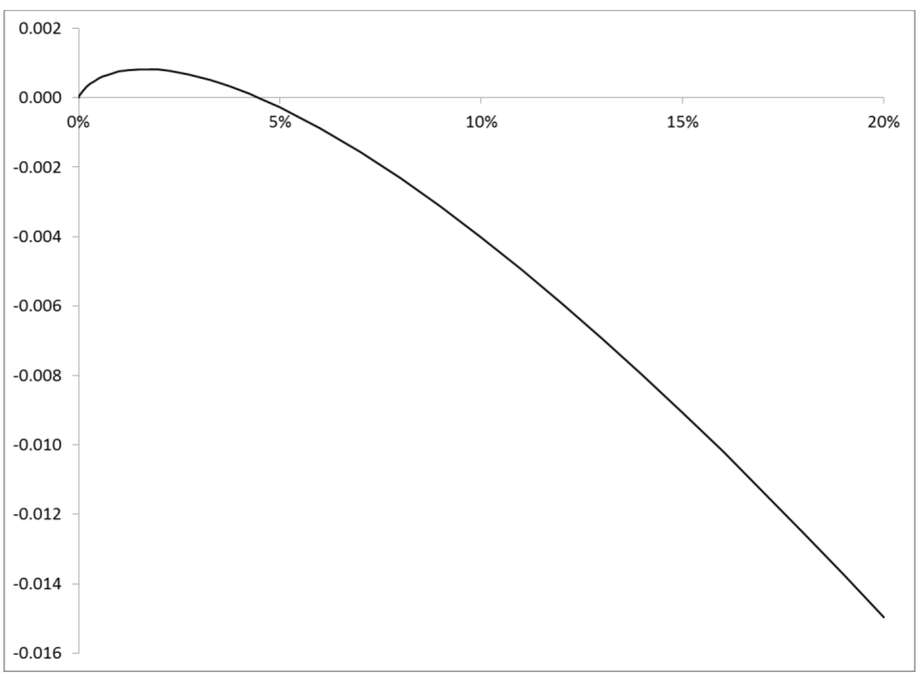

Figure 3

Tree for Example 2

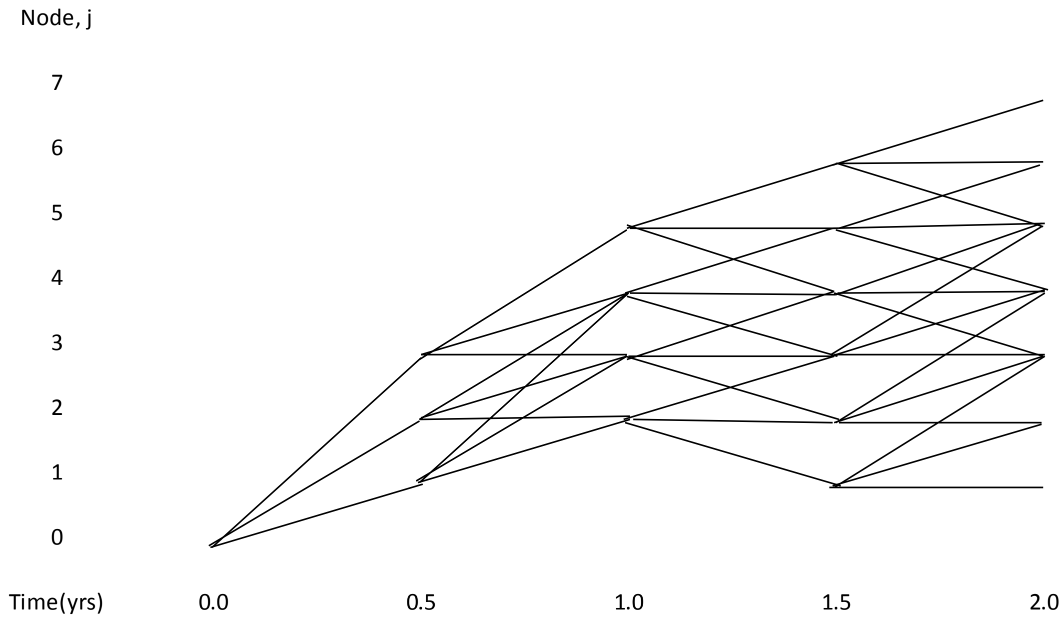

Figure 4

Volatility function in equation (4) used to test model convergence.

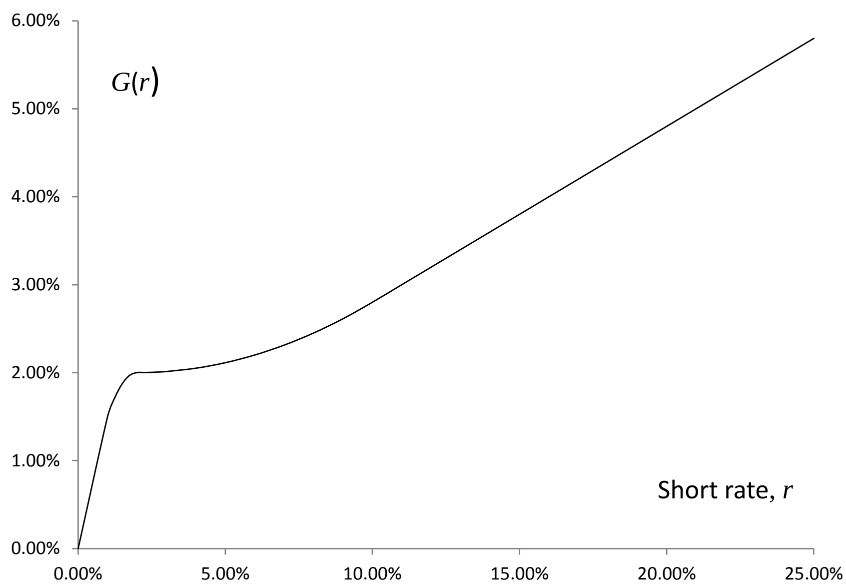

Figure 5:

The best-fit $G(r)$ function calibrated to 10-year cap prices from December 2, 2013. Two different functional forms are used. In the first case $G(r)$ is a piecewise linear function defined by 7 corner points. In the second case $G(r)$ is defined by 3 corner points.

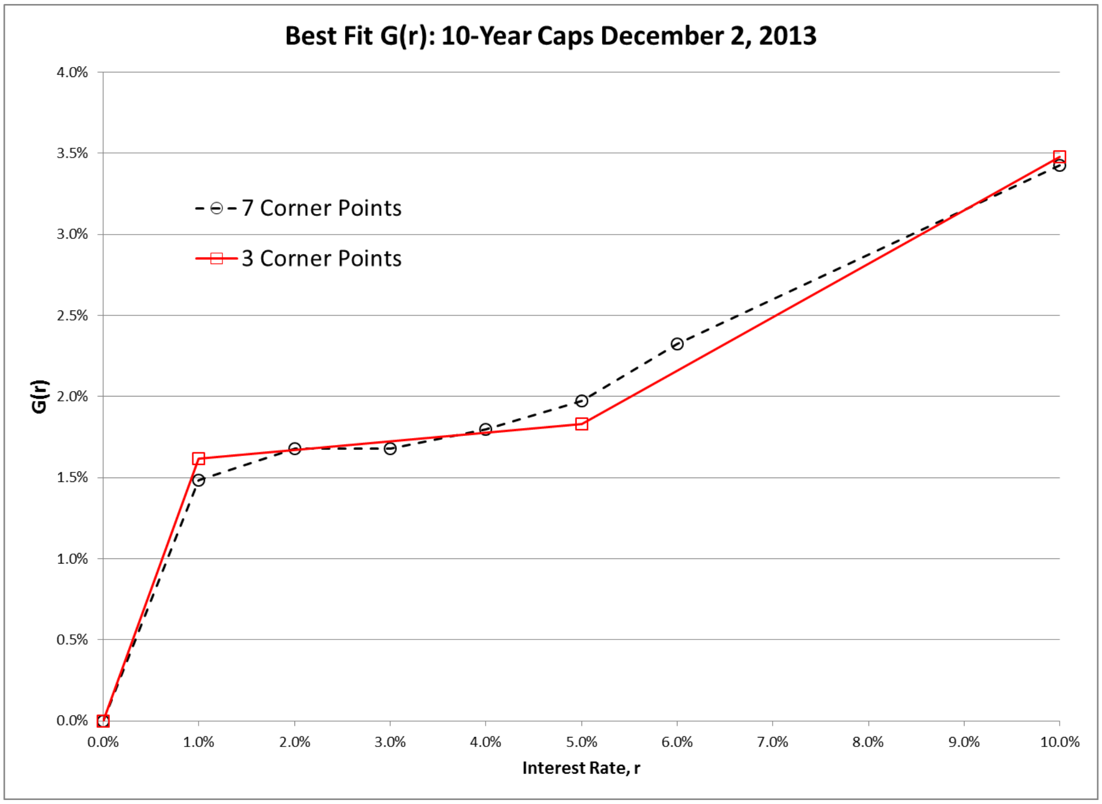

Figure 6:

Deguillaume *et al*'s result for the variability of the short rate as a function of the short rate

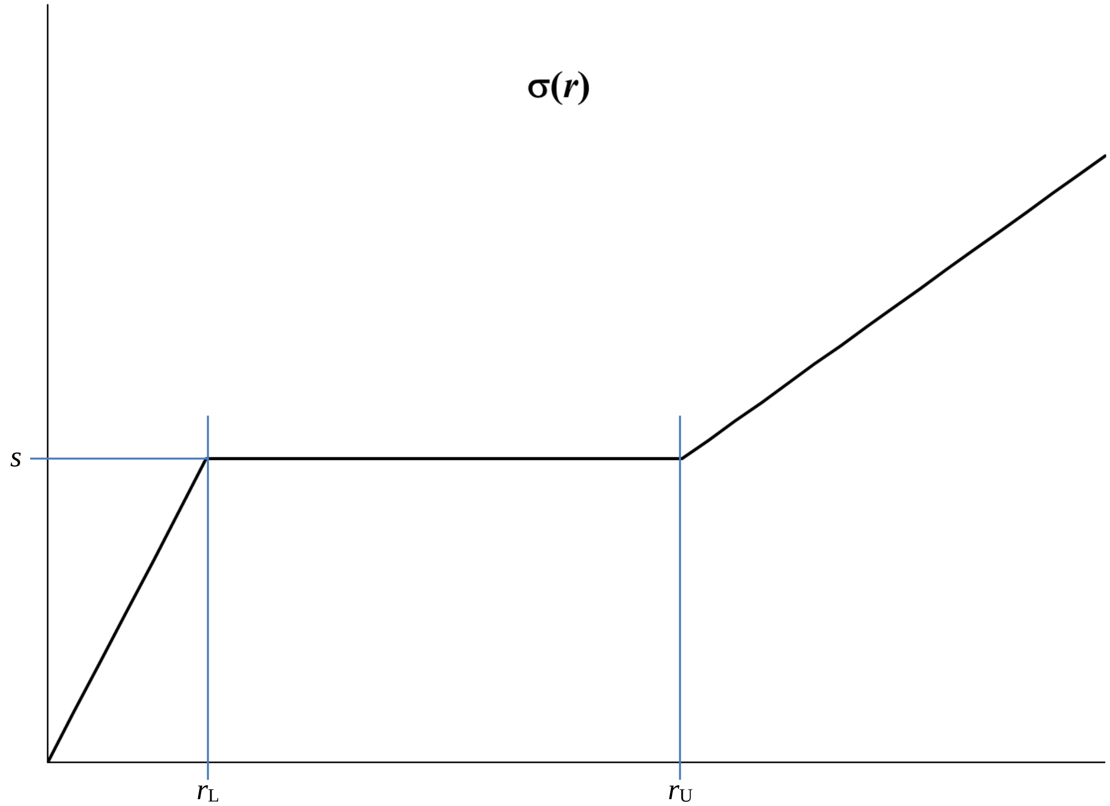

Figure 7

The best-fit $G(r)$ function calibrated to 1- to 10-year cap prices from 2004 to 2013. The solid line shows the average of the ten $G(r)$ functions

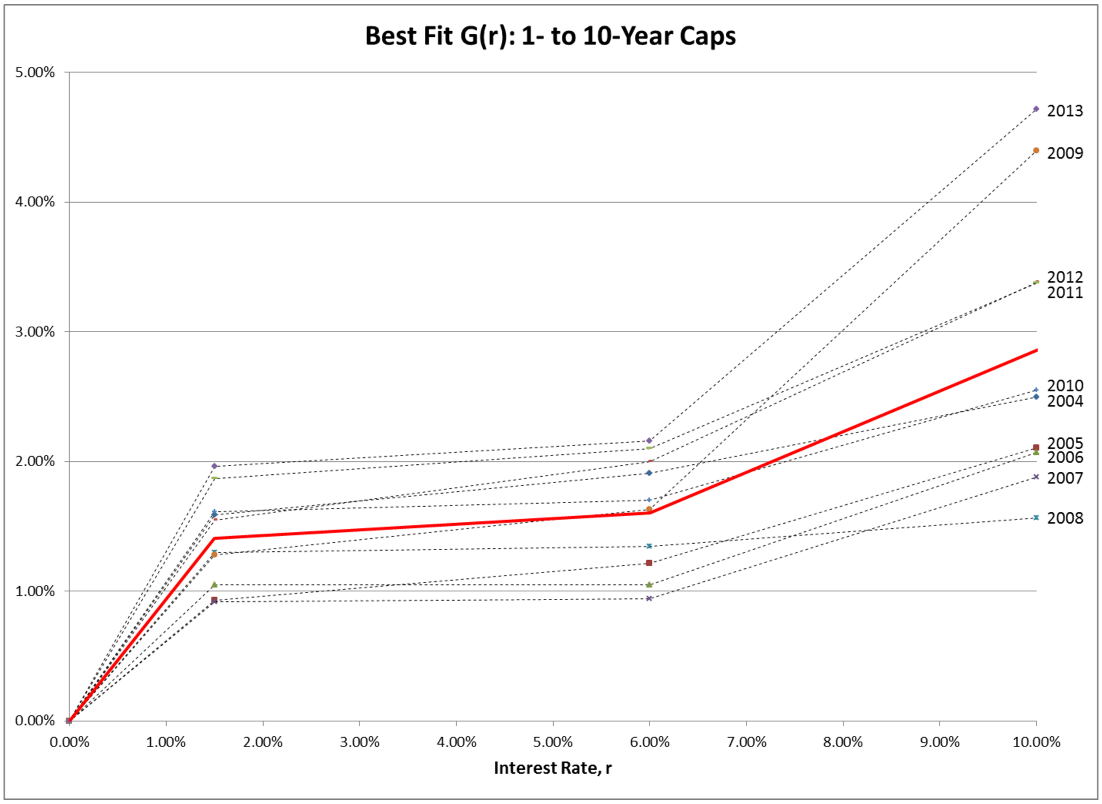

Figure 8

The best-fit $G(r)$ function calibrated to 10- to 30-year cap prices from 2004 to 2013. The solid line shows the average of the ten $G(r)$ functions