RISK SENSITIVITIES OF BERMUDA SWAPTIONS

VLADIMIR V. PITERBARG

Abstract. We present new theoretical results for risk sensitivities of Bermuda swaptions, and derive new representations for them. We apply these results to the problem of risk sensitivities computation and derive algorithms that perform the task much faster and more accurately that the traditional approach. Computation of risk sensitivities to market and model parameters (deltas, gammas, vegas) is one of the most important applications for any model. In most practical situations, the Greeks are computed numerically by shocking appropriate inputs and revaluing the instrument. The time needed to execute such a scheme grows linearly with the number of Greeks required. Our approach allows one to compute any number of Greeks for a Bermuda swaption in nearly constant time. Computational advantages versus the standard approach are significant, with time needed to compute a large number of sensitivities reduced by orders of magnitude. Our approach explores symmetries in the structure of Bermuda swaptions, and is essentially model-independent. The approach is based on a newly discovered set of recursive relations between different sensitivities. The recursive relations allow us to represent sensitivities in a number of interesting ways, in particular as integrals over the "survival" density. The survival density is obtained as a solution to a forward Kolmogorov equation. This representation is the basis for practical applications of our approach.

1. Introduction

Among all exotic interest rate derivatives Bermuda swaptions are by far the most common. Issues related to modeling Bermuda swaptions are undoubtedly a major point of focus for many quantitative research departments. A steady flow of research papers on the subject, and an occasional controversy (see [AA00], [LSSC99]) are a testament to that.

Bermuda swaptions are American-style options on interest rate swaps. An interest rate swap is a contract that specifies an exchange of fixed-rate payments for float-rate payments between two counterparties on a given time schedule. A European swaption is an option to enter the swap on a given date. A Bermuda swaption is a extension of a European swaption. It gives its holder a flexibility to enter the swap on any of the pre-specified dates (instead of just one for a European swaption). If exercised on a given date, the holder enters the part of the swap after the date it was exercised.

Since Bermuda swaptions have an optimal exercise decision feature, they are typically valued by a partial differential equation (PDE) based or a tree based method. A PDE is solved backwards, applying optimal exercise rules on each exercise date.

In practical applications, calculation of risk sensitivities for a Bermuda swaption is as important as calculation of its value. A Bermuda swaption, being an interest-rate derivative,

Date: September 26, 2002.

Key words and phrases. Bermudan swaptions, hedging, Greeks, deltas, vegas, gammas, lattice methods, PDE, BGM, Libor Market Model, Cheyette model.

I would like to thank Paul Cloke, Jesper Andreasen and Leif Andersen for valuable discussions and feedback.

is potentially exposed to movements of each point on the interest rate term curve (up to the underlying swap's final maturity). These risk sensitivities are called deltas. Typically, they are calculated and presented in the following way. A time line is split into time intervals (buckets) of common length, for example three months. Forward interest rates that correspond to all buckets are shocked in turn. A shocked interest rate curve is fed into the model, and the Bermuda swaption is revalued. The sensitivity of the Bermuda swaption value to this shock (often referred to as a "bucketed" delta) is then computed by subtracting a base value of the Bermuda swaption from its shocked value, and normalizing the difference by the size of the bump.

Second-order derivatives to the changes in rates are also common, and usually called bucketed gammas.

Bermuda swaptions are sensitive to changes in interest rate volatility as well. This sensitivity is also usually bucketed, in the sense that the Bermuda swaption sensitivity to various interest rate volatilities is computed. Bermuda swaptions are related to European swaptions. In most markets, European swaptions are actively traded and their implied volatilities are readily observable. For that reason, interest rate volatility sensitivities, called vegas, of Bermuda swaptions are computed with respect to shocks to European swaption volatilities.

Some purists may argue that sensitivities computed by shocking individual pieces of market and model data (various forward rates and volatilities) are inconsistent with the dynamics of the interest rate term curve specified by the model. To them we say that practical considerations involved in managing a an interest rate derivatives book far outweigh theoretical deficiencies of the approach.

The Greeks calculation scheme, as presented above, requires a new value of a Bermuda swaption to be obtained for each bucketed Greek (sometimes even two new valuations are required if two-sided numerical derivatives are used). Thus, computing bucketed deltas to individual shocks of all 3-month tenor rates for a 40 year Bermuda swaption will require about 120 calls to the valuation model. Each valuation requires solving a PDE equation or a rollback on a tree. Adding gammas and vegas increases the number of valuations further. Multiplied by the size of a typical portfolio of Bermuda swaptions, the computational cost adds up pretty quickly.

Another, and perhaps even more important issue with the "shock-and-revalue" approach is related to numerical noise. Each valuation of a Bermuda swaption is computed as a numerical solution to a PDE equation. This introduces numerical errors, or noise. This noise gets multiplied manifold when numerical derivatives of numerical solutions are computed. When second numerical derivatives of numerical solutions are performed, the noise often dominates the value.

Our paper focuses on these issues. It is motivated by a search for "smart" ways to compute Greeks for Bermuda swaptions. We are motivated by the belief that everything that is needed to compute a delta is available when computing a value. For example, this is the case in the Black-Scholes model for options on a stock. In that model a delta (sensitivity to a stock price) can be "read" directly of the PDE grid, from the two lattice grid nodes adjacent to the node where the value is stored.

Another motivation comes from results on computing sensitivities in Monte-Carlo simulation (see [GZ99]). The sensitivities are obtained either by differentiating the payoffs "along the simulated path" or by differentiating the measure under which the value is computed. The results are powerful, and give us incentive to look for similar results in the PDE setting.

Valuation of Bermuda swaptions is based on a set of recursive relations that define the value of a Bermuda swaption with n remaining exercise opportunities as an expected value of the maximum of a Bermuda swaption with $n-1$ remaining exercise opportunities and an exercise value. Starting from these recursive relations for values and "differentiating them through", we derive recursive relations for deltas. These formulas express deltas of a Bermuda swaption with n remaining exercise opportunities as expected values of deltas of a Bermuda swaption with $n-1$ remaining exercise opportunities. The success of this approach relies on defining deltas as sensitivities to specially chosen set of variables that are specific to each Bermuda swaption. In particular, we apply shocks to exercise values directly and individually. Deltas defined as sensitivities to these direct shocks to exercise values allow for especially elegant recursive formulas. At the end, we relate shocks to the initial interest rate curve to these shocks of exercise values, and derive formulas for deltas defined in a standard way.

Similar recursions are derived for gammas and vegas, although under more restrictive conditions.

The recursive relations for Greeks allow us to interpret them in a number of financially meaningful ways. We represent deltas as functionals of the optimal exercise time. We represent them as prices of knock-out contingent claims. We also represent them as integrals of certain payoffs against what we call a "survival" density. The survival density represents the density of the state variables of the model at each point in time given the Bermuda swaption was not exercised up to that time. Financially, the survival density represents a price of a contingent claim paying \$1 in a given state of the world under the condition that the Bermuda swaption was not exercised up until that time.

The latter representation proves the most fruitful. The survival density can be obtained as a solution of a "forward" PDE. It only needs to be computed once. Once it is computed and stored, all Greeks are obtained by simple numerical integration of known payoffs against that density.

Our scheme for computing Greeks brings both speed and accuracy benefits.

The speed benefits of our scheme arise from the fact that, once a survival density is computed, computing each Greek requires only a numerical integration with respect to that density, an essentially "free" operation (compared with a full-blown PDE solution required by the standard "shocking" method).Even more significant savings are realized if other Greeks (gammas and vegas) are requested.

Our scheme not only computes Greeks faster, it often does it more accurately as well. There are two reasons for that. First, we replace the problem of numerical differentiation of numerical solutions to PDE's with that of integration, which is inherently less noisy numerically. Second, because all Greeks are computed from the same survival density, we only have one computation that we need to focus on when trying to increase numerical accuracy. For example, we can decide to use a finer PDE grid for computing the survival density, or implement various smoothing techniques for it.

Our research has both theoretical and practical implications. On the theoretical level, this paper contributes to deeper understanding of Bermuda swaptions and, by extension, American options in general. Our theoretical contributions include obtaining new recursive relationships between different bucketed Greeks, and representing the Greeks in a number of financially meaningful ways. We demonstrate that all risk sensitivities are related to the same object, the survival density. We also derive a PDE for the survival density.

The main implications of our work, however, are deeply practical. Theoretical results we obtain lend themselves very well to practical implementations. Implementing our scheme is no more difficult than implementing a Bermuda swaption valuation method, and its speed and accuracy benefits are hard to ignore.

Our results are obtained with the goal of optimizing Greeks computations in a PDE-based method in mind. However, they prove very useful in a Monte-Carlo setting as well. We derive simple formulas for obtaining Greeks of Bermuda swaptions in the same Monte-Carlo simulation as the values are obtained. The formulas explore the specific structure of Bermuda swaptions and are based on the same Greek representation result that our PDE-based method employs. In this direction, our work extends the work of[GZ99] and specializes it for the case of Bermuda swaptions.

The paper is organized as follows. In Section 2 we place our results in the context of available research. In Section 3 we specify our modeling framework. In the section after that, Section 4, we formally define interest rate instruments that we work with, namely swaps and Bermuda swaptions. We then present a high-level overview of our method in Section 5. We formally define instrument-specific, or "model" deltas in Section 6. We derive our main recursive formulas for "model" deltas in Section 7. Delta representation results are given in Section 8. They are represented as functionals of the optimal exercise time, as knock-out contingent claim values, and as integrals against the survival density. A partial differential equation for the survival density is derived in Section 9. The results of the previous sections are combined in Section 10 where formulas for risk sensitivities (deltas) of Bermuda swaptions with respect to shocks to the initial interest rate curve are presented. We discuss possible further reduction in computational effort in that section by reorganizing the order of computations to reduce the number of integrations performed. The application of our method to gammas (second-order sensitivities to interest rate shocks) are given in Section 11. In Section 12 we present the extension of our method to vega (sensitivity to interest rate volatilities) computations. We discuss what extra conditions on the model are needed for our method to work for vegas, what the right "basis" for computing vega is, and how to translate internal model vegas to market vegas. In Section 13 we extend our results for deltas to models in which volatility depends on interest rates. The extension is straightforward but requires some mild assumptions imposed on the model's volatility process. A short discussion on the distinction between deltas while keeping market volatilities constant versus while keeping model volatilities constant is presented in Section 14. We focus on a particular model next. A case study of the Cheyette model is presented in Section 15. Explicit formulas for deltas, the survival density, the vegas, and volatility correction for the deltas are derived. The results on delta representation are applied to computing Greeks in Monte-Carlo simulation in Section 16, where we present a simple algorithm for computing all deltas of a Bermuda swaption in the same simulation as its value. In Appendix A, we present proofs of various results, and in Appendix B we present recursive formula for bucketed gammas.

2. Literature review

The subject of Bermuda swaption sensitivities has not been extensively treated in the literature. A number of research directions, however, are close in spirit. The paper [GZ99] discusses different way of computing sensitivities of interest rate derivatives in a Monte-Carlo simulation. The authors present two methods, one based on differentiating the measure under which expectations are taken, and the other is based on differentiating the payoffs under the expectation. They generally find that the latter works better. Our method is also based on differentiating payoffs under the expectation operator. However, we take this technique much further by exploring special recursive relations available for Bermuda swaptions. In this case, the formulas lead to representations of sensitivities as integrals against a survival density, a result not available for general interest rate derivatives.

Alternative approaches to computing deltas in Monte-Carlo based on Malliavin calculus are explored in [FLLT99], [Ber99].

Another paper to which ours is related is [Car01]. In it the author finds representations of European option sensitivities (first- and higher-order) as prices of certain contingent claims. Our results extend his to Bermuda swaptions, providing representations of sensitivities of Bermuda swaptions as prices of *knock-out* contingent claims.

Numerical applications of our method come from representing risk sensitivities of a Bermuda swaption in terms of its survival density. The survival density is obtained as a solution to a forward PDE. Related work in [And96] demonstrates the use of forward PDEs in computing risk sensitivities of European options.

Our paper focuses on deriving a method for fast and accurate computation of risk sensitivities for a particular class of interest rate derivatives (Bermuda swaptions). There is a wide body of research available that deals with improving accuracy and speed of tree- and PDE-based methods. In some cases the focus of research is on specific kinds of instruments (for example, barrier options), see [RS97], [RR], [BL94]. In other cases the results and techniques discussed are generic, see [Jam], [RE98], [DJ97], [MM94].

The primary focus of our investigation, Bermuda swaptions, are well covered in the literature. The most popular topics address questions like proper model calibration for a Bermuda swaption ([And01a]); whether or not a multi-factor model is needed to price Bermuda swaptions ([AA00], [LSSC99]); how to value Bermuda swaption in BGM-type models ([And99]), and in general by Monte-Carlo simulation ([LS98]); and what the proper volatility hedges for Bermuda swaptions are ([Dod02]).

Most of the research in interest rate modeling has direct implications for Bermuda swaptions, and is sometimes motivated by applications to Bermuda swaptions. The literature on interest rate modeling is immense; in addition to the papers cited above, [And01b], [AA98], [ABR01] and [JK98] are good indicators of recent trends in this domain.

3. Modeling framework and notations

We start by defining a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ where $\Omega = \{\omega\}$ is a set of elementary events, $\mathcal{F}$ is a sigma-algebra of events, and $\mathbf{P}$ is a probability measure. We assume that an interest-rate model of Heath-Jarrow-Morton type is specified (our approach is largely model independent – we choose to work in the HJM framework for concreteness). In particular, let $\{W(t), t \geq 0\}$ be a K-dimensional Brownian motion. Let $\{\mathcal{H}_t, t \geq 0\}$ be a filtration of sigma-algebras generated by the Brownian motion $\{W(t), t \geq 0\}$ and properly augmented with zero-probability events. Let $C\left([a,b], \mathbb{R}^K\right)$ denote a space of all continuous functions on an interval $[a,b]$ with values in $\mathbb{R}^K$. Let $\sigma(t,T) = \sigma(\omega, t, T)$ be an $\mathcal{H}_t$-adapted stochastic process with values in $C\left([0,\infty], \mathbb{R}^K\right)$, satisfying the usual regularity conditions (see [MR97, Chapter 13]). We write the model for zero-coupon discount bonds $P(t,T) = P(\omega, t, T)$,

$0 \le t \le T < \infty$, as

$$\begin{aligned} dP(t,T) &= r(t) P(t,T)\, dt + P(t,T)\, \sigma(t,T)\, dW(t), \\ P(0,T) &= P_0(T), \quad T \ge 0. \end{aligned} \tag{3.1}$$

Here $r(t)$ is the short rate, and $P_0(\cdot)$ are the initial values of zero-coupon discount bonds, i.e. the interest rate curve as observed today, at time 0. We also define a numeraire, a money-market account

$$B_t = \exp\left(\int_0^t r(s)\, ds\right).$$

The probability measure $\mathbf{P}$ is the risk-neutral probability. We denote the corresponding expectation operator by $\mathbf{E}$ and their $\mathcal{H}_t$-conditioned counterparts by $\mathbf{P}_t$ and $\mathbf{E}_t$.

For each t we define a sigma-algebra $\mathcal{F}_t$ of events generated by bonds observed at time t,

$$\mathcal{F}_t = \sigma\left\{P(t,T), \quad T \ge t\right\}.$$

We define a pricing operator $\pi_t(\cdot)$ for $\mathcal{F}_T$-measurable random variables by

$$\pi_t(\xi) = B_t \mathbf{E}_t B_T^{-1}(\xi).$$

Here $\pi_t(\xi)$ is interpreted as the value, at time t, of a contingent claim with a payoff ξ at time T.

In the measure where $P(\cdot, T)$ is used as a numeraire, the so-called T-forward measure (see [MR97, Chapter 13]), the expression for π becomes

$$\pi_t(\xi) = P(t,T)\, \mathbf{E}_t^T(\xi). \tag{3.2}$$

Here $\mathbf{E}^T$ is the expectation operator under the T-forward measure.

Many of our results are independent of the particular model used. We try to present them in utmost generality. Specific applications, however, require more structure from the model. When necessary or convenient, we will assume that the model is Markov. We call a model Markov if all zero coupon discount bonds can be represented as deterministic functions of a Markov process.

For Markov models, $\pi_t(\xi)$ can be computed as a solution to the model's backward Kolmogorov equation. This technique is especially effective if the dimension of the Markov process is low. Popular choices of Markov models include the Hull-White model ([MR97]), the Cheyette model ([Che91], [And01b]), Kennedy-Hunt family of Markov models ([JK98]), and many others.

In case we want to use or emphasize the Markovian representation of the model we will denote the Markov state process by $x(t) = x(\omega, t)$ and the state space by X $(x(t,\omega) \in X$ for all ω and $t)$. We will assume that all zero coupon discount bonds at time t are deterministic functions of $x(t)$.

Note that in the Markovian setting,

$$\mathcal{F}_t = \sigma\left\{x(t)\right\}.$$

We will generally do not distinguish between $\mathcal{F}_t$-measurable random variables and their representations as deterministic functions of $x(t)$. If we want to emphasize the latter representation we will use x as an argument, i.e. $P(t,T) = P(x(t), t, T)$ where $P(x,t,T)$ is a deterministic function that relates zero coupon bond prices and values of the state process $x(t)$.

We assume that X is a complete metric space. In all practical applications X can be thought of as being a subspace of $\mathbb{R}^M$ for some $M \geq 1$.

We assume that the trajectories of $\mathrm{x}(\cdot)$ are continuous. We denote the Markovian semigroup infinitesimal generator by Λ. The generator Λ is defined on a subspace of the space of all bounded continuous functions on X. In particular, if $v = v(t,x)$ is the value of a European-style contingent claim at time t in the state $x(t) = x$, then $v(t,x)$ satisfies the equation

$$v_t(t,x) + (\Lambda v)(t,x) = r(t)\, v(t,x).$$

4. Instruments

Let $T_0, \ldots, T_N$, $T_n \in \mathbb{R}_+$, be an ordered collection of times,

$$0 = T_0 < T_1 < \cdots < T_N.$$

Such a collection is called a "tenor structure". Denote

$$\tau_n = T_n - T_{n-1}, \quad n = 1, \ldots, N.$$

In defining market instruments we ignore some details that are irrelevant for our presentation. A interest rate swap with the tenor structure $\{T_n\}_{n=0}^N$, fixed rate c, $c > 0$, and unit notional is a contract between two counterparties, a *fixed rate payer* and a *fixed rate receiver*. On each of the dates T_n, $n = 1, \ldots, N$, a fixed rate payer pays the amount $c\tau_n$ to the fixed rate receiver. In return, fixed rate payer receives a payment, also on T_n, that is based on a simple floating rate over the previous period. The amount is equal to

$$\frac{1 - P(T_{n-1}, T_n)}{\tau_n P(T_{n-1}, T_n)} \times \tau_n.$$

By $E_n(t)$ we denote the value, from the point of view of fixed rate payer at time t, of all the cashflows in the swap scheduled on or after T_{n+1} :

$$E_n(t) = \sum_{k=n+1}^{N} \pi_t \left(\frac{1 - P(T_{k-1}, T_k)}{\tau_k P(T_{k-1}, T_k)} \times \tau_k - c\tau_k \right).$$

The formula can be simplified to yield

$$E_n(t) = P(t, T_n) - P(t, T_N) - c \sum_{k=n+1}^{N} P(t, T_k)\, \tau_k, \quad n = 1, \ldots, N-1. \tag{4.1}$$

For completeness we set

$$E_N(t) \equiv 0.$$

The value of the swap is just a linear combination of zero-coupon bonds.

The swap as presented here is of the most simple variety. More complex swaps can be specified with fixed rates that are different for different time periods, the notionals to which rate are applied not constant, and so on. Our approach works for all those varieties with virtually no changes, but we present it in the simplest case to avoid cumbersome notations.

A Bermuda swaption gives its holder a right, but not an obligation, to enter the swap on any of the dates T_n, $n = 1, \ldots, N$. The swap that is actually entered to is the part remaining after the exercise date T_n. Once the Bermuda swaption is exercised, it disappears.

A Bermuda swaption is an American-style derivative with multiple (albeit discrete) exercise opportunities.

For future considerations it is important to define a whole family of Bermuda swaptions. By $H_n(t)$ we denote the value, at time t, of a Bermuda swaption that has only the dates $\{T_{n+1}, \ldots, T_{N-1}\}$ as exercise opportunities. In particular, $H_0(0)$ is the value of the Bermuda swaption we re interested in at time zero. Necessarily

$$H_0(t) \geq H_1(t) \geq \cdots \geq H_{N-2}(t).$$

We will speak of a "Bermuda swaption H_n" as a shorthand for "the Bermuda swaption whose value at time t is equal to $H_n(t)$."

4.1. **Recursion for Bermuda swaptions.** If the Bermuda swaption H_0 has not been exercised up to and including time T_n, ("still alive at time T_n") then it is worth exactly the same as the Bermuda swaption H_n (hence 'H' for "hold value"). If the Bermuda swaption is exercised at time T_n its value is equal to $E_n(T_n)$ (hence 'E' for "exercise value"). Assuming optimal exercise, the value of the Bermuda swaption H_0 at time T_n is then the maximum of the two,

$$\max\{H_n(T_n), E_n(T_n)\}.$$

The value of this payoff at time T_{n-1} is then

$$\pi_{T_{n-1}} \max\{H_n(T_n), E_n(T_n)\}.$$

Clearly this is the value of the Bermuda swaption that can only be exercised at time T_n and beyond, i.e. of the Bermuda swaption H_{n-1}. These considerations lead to a recursion

$$\begin{aligned} H_{n-1}(T_{n-1}) &= \pi_{T_{n-1}} \max\{H_n(T_n), E_n(T_n)\}, \quad n = N-1, \ldots, 1, \\ H_{N-1} &\equiv 0. \end{aligned} \tag{4.2}$$

The recursion starts at the final time $n = N-1$ and progresses backward in time. For $n = 1$ we obtain the value $H_0(0)$, the value of the Bermuda swaption that we are after.

This is of course nothing more than a well-known algorithm for pricing American-style options on a grid. This recursion, however, is our starting point in deriving an efficient algorithm for computing Greeks.

5. Roadmap of the method

Before proceeding with the details of our method for efficient computation of all sensitivities for a Bermuda swaption, we would like to give a broad overview of the method, and key insights that led us to it.

Sensitivities to shocks of the interest rate curve are most often computed for a standard set of market "bumps". One popular choice is to use 3-month forward rates every three months, and compute sensitivities (bucketed deltas) of an instrument to independent shocks to these rates.

Our first key insight is that bucketed deltas can be computed in any "basis". Bucketed deltas are first-order derivatives. First-order derivatives behave very well under a change of variables. To compute a vector of derivatives with respect to a set of variables, we can compute a vector of derivatives with respect to another set of variables, and then multiply it by the Jacobian of the variables' change. In short, we can choose what to bump for each Bermudan – we can always "rotate" the exposures into whatever coordinates.

Armed with this insight, as our first step we come up with a set of inputs that we will bump. This set will be specific to each Bermudan.

We take this a step further and define deltas not in terms of the bumps to the initial interest rate curve, but in terms of the bumps to the values of some interest rate instruments *in the future*. For that, we define a Bermuda swaption as a function that maps a collection of random variables representing future values of the underlying to a value. *We differentiate this function.*

On the next step, we derive recursions for the sensitivities as defined above. We use the recursion (4.2) and "differentiate through" it to get the required relations between sensitivities.

With the help of recursive relations obtained, we represent sensitivities in a number of financially-meaningful ways, in particular as integrals with respect to the survival density.

The equation for the survival density is derived next. This equation is central to our goal of reducing the computational cost of computing sensitivities because with that equation, the survival density needs to be computed only once for all sensitivities.

As a final step, the model sensitivities are "rotated" into the sensitivities to moves in the initial interest rate curve.

The same approach is then applied to vegas, the sensitivities to changes in various volatilities.

6. Defining "model" deltas

Let us rewrite (4.2) in a slightly more convenient way. Let us set

$$F(t,T) = \frac{1}{T-t} \log P(t,T). \tag{6.1}$$

Here $F(t,T)$ is a continuously compounded spot rate for the period $[t,T]$ observed at time t. Using forward measures we can rewrite (4.2) as

$$H_{n-1}(T_{n-1}) = P(T_{n-1},T_n)\,\mathbf{E}^{T_n}_{T_{n-1}} \max\{H_n(T_n), E_n(T_n)\}, \quad n = N-1,\ldots,1.$$

Using the definition (6.1) we obtain

$$H_{n-1}(T_{n-1}) = \exp(-\tau_n F(T_{n-1},T_n))\,\mathbf{E}^{T_n}_{T_{n-1}} \max\{H_n(T_n), E_n(T_n)\}, \quad n = N-1,\ldots,1. \tag{6.2}$$

We look at this set of relations as defining an algorithm, or a deterministic function, to obtain the value $H_0(0)$, as well as all intermediate values $H_n(T_n)$, from a collection of inputs. What are the inputs to this function? The inputs are

$$\{E_n(T_n)\}_{n=1}^{N-1}, \tag{6.3}$$

$$\{F(T_{n-1},T_n)\}_{n=1}^{N-1}. \tag{6.4}$$

The inputs on the line (6.3) are the exercise values. The inputs on the line (6.4) are discount rates as observed on exercise dates.

Our method is based on the idea that all quantities in (6.3)-(6.4) should be understood as independent inputs to the "Bermuda pricing function" alluded to earlier. This is not to say that we do not recognize that all of them are deeply related. However, these inputs do enter the "expression" (6.2) separately and therefore, can be shocked individually. This is the key idea of our method.

In due course, of course,the dependencies between these sets of inputs are recognized. These dependencies enter our final formulas that combine the sensitivities to these parameters into sensitivities to the "standard" set of market parameters.

Let us formalize these ideas and rigorously define the derivatives. Let $\mathcal{E}_n$, $0 \leq 1 \leq N-1$, be a Banach space of all $\mathcal{F}_{T_n}$-measurable $\mathbb{R}$-valued random variables equipped with the L_1 norm. Denote

$$\mathcal{A} = \left(\bigoplus_{n=1}^{N-1} \mathcal{E}_n \right) \oplus \left(\bigoplus_{n=0}^{N-2} \mathcal{E}_n \right). \tag{6.5}$$

From now on, we adopt a dual view-point on the equation (6.2). In addition to defining an algorithm to value a Bermuda swaption, we view it as a definition of a functional on the space $\mathcal{A}$ defined in (6.5).We do that in the following sense. Let for any n, e_n and f_n be arbitrary elements of $\mathcal{E}_n$. Define

$$a = (e_1, \ldots, e_{N-1}, f_0, \ldots, f_{N-2}).$$

By definition $a \in \mathcal{A}$. Then for each k, a function

$$h_k : \mathcal{A} \to \mathcal{E}_k$$

is defined recursively by

$$\begin{aligned} h_{k-1}(a) &= \exp(-\tau_k f_{k-1}) \mathbf{E}_{T_{k-1}}^{T_k} \max\{h_k(a), e_k\}, \quad k = N-1, \ldots, 1, \\ h_{N-1} &\equiv 0 \end{aligned} \tag{6.6}$$

(compare this to (6.2)). In this definition, the mathematical expectation operator $\mathbf{E}_{T_{k-1}}^{T_k}$ is viewed as a linear operator between two functional spaces,

$$\mathbf{E}_{T_{k-1}}^{T_k} : \mathcal{E}_k \to \mathcal{E}_{k-1}.$$

It is these functions $h_n(a)$ that we differentiate with respect to various coordinates of a.

Let us denote by A an element of $\mathcal{A}$ that represents the collection of "real" values of the exercise swaps and rates,

$$A = (E_1(T_1), \ldots, E_{N-1}(T_{N-1}), F(T_0, T_1), \ldots, F(T_{N-2}, T_{N-1})). \tag{6.7}$$

Then, comparing (6.2) and (6.6) and using the definition of A, we obtain

$$H_k(T_n) = h_k(A)$$

for each k, $0 \leq k \leq N-1$.

Each coordinate of a, $a \in \mathcal{A}$, is a random variable. We understand a derivative with respect to a random variable as a derivative in a specific "direction", with the direction being a random variable itself. Specifically, for each n, $1 \leq n \leq N-1$, we choose $d_n^u \in \mathcal{E}_n$ and for each n, $0 \leq n \leq N-2$, we choose $d_n^d \in \mathcal{E}_n$. Furthermore, we set $\bar{d}_n^u$ to be the element of $\mathcal{A}$ whose n-th coordinate is d_n^u and all other coordinates are zero, and we set $\bar{d}_n^d$ to be the element of $\mathcal{A}$ whose $(N+n)$-th coordinate is d_n^d and all other coordinates are zero:

$$\bar{d}_n^u = \left(0, \ldots, \underset{\overset{\uparrow}{n}}{d_n^u}, 0, \ldots, \underbrace{0, \ldots, 0}_{N-1}\right),$$

$$\bar{d}_n^d = \left(\underbrace{0, \ldots, 0}_{N-1}, 0, \ldots, \underset{\overset{\uparrow}{N+n}}{d_n^d}, 0, \ldots, 0\right).$$

The derivatives can now be defined. For each h_k, $1 \le k \le N-1$, the *underlying bucketed deltas* are defined to be the derivatives of h_k in the directions $\bar{d}_n^u$,

$$\Delta_n^u h_k(a) \triangleq \lim_{\varepsilon \downarrow 0} \varepsilon^{-1}\left(h_k\left(a + \varepsilon \bar{d}_n^u\right) - h_k(a)\right), \quad k = 1, \ldots, N-1 \quad n = 1, \ldots, N-1.$$

For each h_k, $0 \le k \le N-1$, the *discount bucketed deltas* $\Delta_n^d h_k$ are defined as the derivatives of h_k in the directions $\bar{d}_n^d$, $0 \le n \le N-2$. Specifically,

$$\Delta_n^d h_k(a) \triangleq \lim_{\varepsilon \downarrow 0} \varepsilon^{-1}\left(h_{nk}\left(a + \varepsilon \bar{d}_n^d\right) - h_k(a)\right), \quad k = 0, \ldots, N-1, \quad n = 0, \ldots, N-2.$$

To understand these quantities a bit better, it is illuminating to look at a simple example of the Hull-White model. The Hull-White model is a single-factor Markovian model with the short rate $r(t)$ being the factor (see. [MR97, Chapter 12]). In particular, any $\mathcal{F}_t$-measurable random variable is a deterministic function of $r(t)$. So for each n, $0 \le n \le N-1$, the space $\mathcal{E}_n$ can be mapped to a subspace of a space of measurable $\mathbb{R} \to \mathbb{R}$ functions (functions of the short rate observed at time T_n). Then we can think of, for example, $d_n^u = d_n^u(r)$ as deterministic functions that are "perturbations" to be added to deterministic functions that correspond to the exercise values $e_n = e_n(r)$.

In the next section recursive formulas for the deltas $\Delta_k^u h_n$ and $\Delta_k^d h_n(a)$ *.are* derived These deltas of course depend on the directions d_n^u and d_n^d. At this point the choice of directions is irrelevant.

7. Recursive formulas for deltas

The derivation of recursive formulas for Greeks is based on the following lemma (proved in Appendix A).

Lemma 7.1. *Let X, Y and D be three $\mathcal{F}_T$-measurable random variables such that*

$$\mathbf{E}|X| < \infty,$$
$$\mathbf{E}|Y| < \infty,$$
$$\mathbf{E}|D| < \infty.$$

Assume that $\mathbf{P}(X = Y) = 0$. *Then*

$$\lim_{\varepsilon \to 0} \varepsilon^{-1} \mathbf{E}_t\left(\max(X + \varepsilon D, Y) - \max(X, Y)\right) = \mathbf{E}_t\left(1_{\{X > Y\}} D\right). \tag{7.1}$$

The expression on the right hand side of (7.1) is a derivative of the function $\max(\cdot, Y)$ in the "direction" D. The importance of this lemma is that the order of differentiation and expectation, even for non-differentiable payoffs such as in (7.1), can be exchanged for such directional derivatives.

The ability to differentiate such non-smooth payoffs hinges on the fact that the expectation operator is "smoothing".

The following theorem is the first main result of this section.

Theorem 7.2. *Recall the definition of* A *in (6.7). Let* k, n *be such that* $0 \le k \le N-1$, $1 \le n \le N-1$. *Assume that*

$$\mathbf{P}\left(H_n\left(T_n\right) = E_n\left(T_n\right)\right) = 0, \quad n = 1, \dots, N-1.$$

Then the following holds for the underlying bucketed deltas.

1. For $k > n-1$,
$$\Delta_n^u h_k\left(A\right) = 0;$$
2. For $k = n-1$,
$$\Delta_n^u h_k\left(A\right) = \pi_{T_k}\left(1_{\{H_n(T_n) \le E_n(T_n)\}} \times d_n^u\right);$$
3. For $k < n-1$,
$$\Delta_n^u h_k\left(A\right) = \pi_{T_k}\left(1_{\{H_{k+1}(T_{k+1}) > E_{k+1}(T_{k+1})\}} \times \Delta_n^u h_{k+1}\left(A\right)\right).$$

In particular, for $1 \le n \le N-1$,

$$\Delta_n^u h_0\left(A\right) = \pi_0\left(\prod_{j=1}^{n-1} 1_{\{H_j(T_j) > E_j(T_j)\}} \times 1_{\{H_n(T_n) \le E_n(T_n)\}} \times d_n^u\right). \tag{7.2}$$

An equivalent result holds for discount deltas:

Theorem 7.3. *Let* k, n *be such that* $0 \le k \le N-1, \quad 0 \le n \le N-2$. *Assume that*

$$\mathbf{P}\left(H_n\left(T_n\right) = E_n\left(T_n\right)\right) = 0, \quad n = 1, \dots, N-1.$$

Then the following holds for the discount bucketed deltas

1. For $k > n$,
$$\Delta_n^d h_k\left(A\right) = 0;$$
2. For $k = n$,
$$\Delta_n^d h_k\left(A\right) = -\tau_{n+1} \times d_n^d \times H_k\left(T_k\right);$$
3. For $k < n$,
$$\Delta_n^d h_k\left(A\right) = \pi_{T_k}\left(1_{\{H_{k+1}(T_{k+1}) > E_{k+1}(T_{k+1})\}} \times \Delta_n^d h_{k+1}\left(A\right)\right).$$

In particular, for $0 \le n \le N-2$,

$$\Delta_n^d h_0\left(A\right) = -\tau_{n+1}\pi_0\left(\prod_{j=1}^{n-1} 1_{\{H_j(T_j) > E_j(T_j)\}} \times 1_{\{H_n(T_n) > E_n(T_n)\}} \times d_n^d \times H_n\right). \tag{7.3}$$

Remark 7.1. *The condition*

$$\mathbf{P}\left(H_n\left(T_n\right) = E_n\left(T_n\right)\right) = 0, \quad n = 1, \dots, N-1,$$

imposed by the two theorems is purely technical. Only highly degenerate models do not satisfy it.

These two theorems define recursive formulas to compute the underlying and discount bucketed deltas.

The proof of these two theorems is given in the Appendix A. It is based on the repeated application of Lemma 7.1. We “differentiate through” the recursion for the values (6.6) to obtain the recursion for deltas.

8. Financial interpretation of Bermuda swaption deltas

In this section we provide different interpretations of deltas of a Bermuda swaption that stem from the results in Theorems 7.2 and 7.3. We demonstrate the connection between deltas and the optimal exercise time for a Bermuda swaption; deltas and prices of knock-out contingent claims; and deltas and the survival measure. It is the last connection that turns out to be the most fruitful. We explore it in subsequent sections.

8.1. **Deltas and optimal exercise time.** Let us define the optimal exercise index for the Bermuda swaption by η, i.e.

$$\eta = \min\left\{n \geq 1 : H_n\left(T_n\right) \leq E_n\left(T_n\right)\right\}.$$

The value of the Bermuda swaption can be written in terms of the optimal exercise time as follows,

$$\begin{aligned} h_0\left(A\right) &= \sum_{m=1}^{N-1} \pi_0\left(1_{\{\eta=m\}} E_m\left(T_m\right)\right) \\ &= \pi_0\left(E_{\eta\wedge N}\left(T_{\eta\wedge N}\right)\right). \end{aligned} \tag{8.1}$$

The following corollary shows that the deltas can be written in terms of the optimal exercise time as well.

Corollary 8.1. *The following holds for any* n, $1 \leq n \leq N-1$,

$$\begin{gathered} \Delta_n^u h_0\left(A\right) = \pi_0\left(1_{\{\eta=n\}} d_n^u\right), \\ \Delta_n^d h_0\left(A\right) = -\tau_{n+1} \pi_0\left(1_{\{\eta>n\}} d_n^d H_n\left(T_n\right)\right). \end{gathered} \tag{8.2}$$

Proof. The result follows from formulas (7.2) and (7.3), and the representations

$$\begin{aligned} 1_{\{\eta=n\}} &= \prod_{j=1}^{n-1} 1_{\{H_j(T_j)>E_j(T_j)\}} \times 1_{\{H_n(T_n)\leq E_n(T_n)\}}, \\ 1_{\{\eta>n\}} &= \prod_{j=1}^{n-1} 1_{\{H_j(T_j)>E_j(T_j)\}} \times 1_{\{H_n(T_n)>E_n(T_n)\}}. \end{aligned}$$

∎

The corollary states that the derivative of the Bermuda value with respect to a directional shock to the n-th exercise value is a discounted expected value of that shock restricted to the event that Bermuda swaption was exercised exactly on the n-th exercise opportunity.

Note that the statement (8.2) can be obtained from (8.1) by formally differentiating it with respect to $E_n\left(T_n\right)$ and assuming that η does not depend on $E_n\left(T_n\right)$.

Corollary 8.1 will provide a basis for computing deltas of a Bermuda swaption in a Monte-Carlo simulation. This direction will be explored later in the paper.

8.2. **Deltas and knock-out options.** The corollary below exposes an interesting and intriguing connection between deltas of Bermuda swaptions and values of knock-out options. Recall that we denoted by X the state space of the Markov state process $x\left(t\right)$ and the notation $H_n\left(T_n\right) = H_n\left(x\left(T_n\right), T_n\right)$, etc. Let us define the n-th exercise region by

$$R_n = \left\{x \in X : H_n\left(x, T_n\right) \leq E_n\left(x, T_n\right)\right\} \subset X, \quad 0 \leq n \leq N-1.$$

We call a contingent claim with an $\mathcal{F}_{T_n}$-measurable payoff ξ a T_n-*knock-out* if the contingent claim disappears when on any of the times $T_1, \ldots, T_{n-1}$, the state process $x(T_i)$ enters the exercise region R_i.

The following statement is another interpretation of the formulas (7.2) and (7.3).

Corollary 8.2. *The n-th underlying delta, $\Delta_n^u h_0(A)$, is equal to the value of a T_n-knockout contingent claim with the payoff $1_{\{H_n(T_n) \leq E_n(T_n)\}} \times d_n^u$. The n-th discount delta, $\Delta_n^d h_0(A)$, is equal to the value of a T_n-knockout contingent claim with the payoff $-\tau_{n+1} 1_{\{H_n(T_n) > E_n(T_n)\}} \times d_n^d \times H_n$.*

Carr in [Car01] provided the interpretation of Greeks of European options as values of certain contingent claims. Corollary 8.2 extends this interpretation to Bermuda-style derivatives. Deltas of Bermuda swaptions can be represented as values of *knock-out* contingent claims.

8.3. **Deltas and the survival measure.** Let $n(t)$ be defined as follows,

$$n(t) = \max\{n : T_n < t\}.$$

The index $n(t)$ is uniquely defined by the property that

$$T_{n(t)} < t \leq T_{n(t)+1}.$$

More structure is imposed on the state space $X.in$ this section It is assumed that X has a canonical measure ν, and the Borel sigma-algebra of measurable subsets of X is denoted by $\mathcal{B}(X)$. (In all situations of practical interest, X can be taken to be a subset of $\mathbb{R}^M$ for some M, with ν being the Lebesgue measure.) The time-t survival measure $\Psi(\cdot; t)$ is defined on the state space X by the formula

$$\Psi(Y; t) = \pi_0\left(1_{\{\eta > n(t)\}} \times 1_{\{x(t) \in Y\}}\right),$$

for

$$Y \in \mathcal{B}(X).$$

Here, as before, η is the optimal exercise time index. In the terminology of the previous section, $\Psi(Y; t)$ is the value of the t-knock-out option with the payoff $1_{\{x(t) \in Y\}}$.

The survival density, the density $\psi(x; t)$ of the survival measure with respect to the canonical one ν is defined via the equation

$$\Psi(Y; t) = \int_Y \psi(y; t)\ \nu(dy).$$

It is assumed to exist.

Since the model is Markovian with $x(t)$ being the state process, the "direction" d_n^u is a deterministic function of the state process $x(\cdot)$ evaluated at time T_n,

$$d_n^u = d_n^u(x(T_n)),$$

(we denote the deterministic function by the same symbol to avoid needless clutter of notations). The following corollary holds.

Corollary 8.3. *For each n, $1 \leq n \leq N-1$, the n-th underlying delta can be represented as an integral of a payoff with respect to the survival measure,*

$$\Delta_n^u h_0(A) = \int_X 1_{\{R_n\}} d_n^u(x)\ \Psi(dx; T_n).$$

Similar result holds for the n-th discount deltas. For n, $0 \le n \le N-2$,

$$\Delta_n^d h_0(A) = -\tau_{n+1} \int_X 1_{\{R_n^c\}} d_n^d(x) H_n(x, T_n) \Psi(dx; T_n).$$

These formulas can be written in terms of the survival density as well,

$$\Delta_n^u h_0(A) = \int_X 1_{\{R_n\}} d_n^u(x) \psi(x; T_n) \nu(dx),$$

$$\Delta_n^d h_0(A) = -\tau_{n+1} \int_X 1_{\{R_n^c\}} d_n^d(x) H_n(x, T_n) \psi(x; T_n) \nu(dx).$$

We call $\Psi(\cdot; t)$ the "survival measure" because it assigns values to subsets of X given that the Bermudan swaption did not get exercised ("survived") until time t.

9. Computing survival density on a lattice

In this section we construct an efficient numerical procedure for computing the survival density on a PDE lattice. Before we discuss the computation of the survival density, let us look at the following family of measures defined on the same Borel sigma-algebra $\mathcal{B}(X)$. We fix time s and position x and define

$$\Phi_{s,x}(Y; t) \triangleq \pi_s\left(1_{\{x(t)\in Y\}}\right)(x),$$

(the value of the contingent claim that pays $1_{\{x(t)\in Y\}}$ evaluated at time s given that $x(s) = x$). We denote by $\phi_{s,x}(y; t)$ its density with respect to the canonical measure ν,

$$\Phi_{s,x}(Y; t) = \int_Y \phi_{s,x}(y; t) \nu(dy)$$

(assuming it exists).

We have

$$\begin{aligned} \Phi_{s,x}(Y; t) &= \mathbf{E}_{s,x}\left(B_t^{-1} 1_{\{x(t)\in Y\}}\right) \\ &= \mathbf{E}_{s,x}\left(e^{-\int_s^t r(u)\,du} 1_{\{x(t)\in Y\}}\right). \end{aligned}$$

It is known that the transition density of the Markov process x satisfies the forward Kolmogorov equation. The difference between ϕ and the transition density is the presence of the term $e^{-\int_0^t r(u)\,du}$. A simple extension of Kolmogorov's forward equation leads to the following result.

Proposition 9.1. *For each s, x, the density $\phi_{s,x}(y; t)$ satisfies the forward equation*

$$\frac{\partial}{\partial t}\phi_{s,x}(y; t) = \left(\Lambda^* \phi_{s,x}\right)(y; t) - r(t)\,\phi_{s,x}(y; t) \tag{9.1}$$

for $t \ge s$. Here Λ^* is the operator adjoint to Λ, the generator of the Markov semi-group for the process $x(\cdot)$. It is applied to $\phi_{s,x}(t, y)$ as a function of y.

The following theorem outlines an efficient procedure for computing the survival density ψ. The idea of the theorem is that in between the "interesting" times $\{T_n\}_{n=0}^N$, the density ψ behaves just like the density ϕ in the proposition above. When the time crosses an exercise time T_n, the density ψ gets multiplied by an extra "survival" indicator function $1_{\{y\in R_n^c\}}$.

Theorem 9.2. *For each n, $0 \leq n \leq N-1$, the survival density $\psi(y;t)$ satisfies the forward PDE*

$$\frac{\partial}{\partial t}\psi(y;t) = (\Lambda^* \psi)(y;t) - r(t)\,\psi(y;t),$$

for

$$t \in (T_n, T_{n+1}],$$

with the initial condition

$$\psi(y; T_n + 0) = \psi(y; T_n) \times 1_{\{y \in R_n^c\}}. \tag{9.2}$$

The initial condition for the first interval, $[T_0, T_1]$, is given by the delta function

$$\psi(y; T_0) = \delta_{x(0)}(y).$$

The proof of the theorem is in Appendix A.

The time-T_n conditions (9.2) require knowledge of the "hold" regions

$$R_n^c = \{x \in X : H_n(x, T_n) > E_n(x, T_n)\}.$$

These are computed as a by-product of the Bermuda swaption valuation, since on each exercise date T_n, the hold values $H_n(x, T_n)$ are determined as functions of the state process $x(\cdot)$ evaluated at time T_n.

The theorem outlines a procedure for computing the survival density in one forward PDE "sweep". We start at time zero with a delta function. We roll it forward on our PDE lattice until the first exercise time T_1. At this point we multiply the density by the indicator function of no-exercise condition. We roll it forward until the next exercise date where we multiply it by another no-exercise indicator function, and so on.

This procedure is similar to computing European option prices and deltas in forward PDE (Andreasen scheme in Black-Scholes framework, see [And96]). The extra twist that the Bermuda swaptions bring is the necessity of multiplying the solution by certain indicator functions on exercise dates.

10. Deltas to shocks of the initial interest rate curve

So far the machinery for computing deltas with respect to "directional" shocks of (random) exercise values was developed. Our achievements are impressive and are worth recapping. Each underlying and discount delta has been represented as an integral of a certain payoff with respect to the survival density (Corollary 8.3). An efficient forward-PDE based algorithm for the survival density has been developed (Theorem 9.2). It has been demonstrated that all underlying and discount deltas can be computed by solving a single PDE forward for the survival density (an operation roughly equivalent in terms of computation complexity to a single valuation of a Bermuda swaption), and then $2(N-1)$ integrations that are cheap numerically. However, this is not the end of the story. What is really needed are deltas with respect to specific shocks of the initial interest rate curve, and not deltas to exercise and discount random variables. This section outlines how to apply the machinery developed so far to this practical problem.

An interest rate curve can be parametrized in a number of different ways. Popular choices include parametrizing it in terms of zero-coupon discount bonds, instantaneous forward rates, and so on. All these representations are equivalent. Shocks to the initial interest rate curve, accordingly, can also be expressed in a number of equivalent ways (i.e. shocks to the bonds, shocks to forward rates, and so on). Since our model (3.1) is written in the zero-coupon discount bond terms, we will think of the shocks in terms of the shocks to the function

$$\begin{aligned} P_0(\cdot) &: \mathbb{R}_+ \to (0,1), \\ P_0(\cdot) &: t \mapsto P(0,t), \end{aligned}$$

that maps maturity time to a zero coupon bond value for that maturity. Popular style of shocking interest rate curves, such as shocking consecutive Libor rates with 3-month or 6-month tenors, or shocking market instruments (spot rates, futures and swaps) that go into curve construction can be easily mapped to the shocks to zero coupon bonds.

10.1. **Single shock.** Let $\theta(\cdot)$ be an arbitrary bounded continuous function on $\mathbb{R}_+$ with $\theta(0) = 0$. We will interpret θ as a shock to be applied to the initial interest rate curve expressed in zero coupon bond terms. Define a perturbed initial interest rate curve by

$$P_0^\varepsilon(t) = P_0(t) + \varepsilon\theta(t).$$

We assume that there exists ε_0 such that for $\varepsilon \in [0, \varepsilon_0]$, $P_0^\varepsilon(t)$ is a valid interest rate curve (i.e. $P_0^\varepsilon(t) \in (0,1)$ for $t > 0$ and it is monotonically decreasing in t).

We would like to consider an interest rate model that is the same as (3.1) but uses $P_0^\varepsilon(\cdot)$, instead of $P_0(\cdot)$, for the initial values in the evolution of zero-coupon discount bonds. The volatility process $\sigma(t,T)$, as defined, can be a function of the initial interest rate curve as well.

For the duration of this section we assume that the volatility function $\sigma(t,T)$ is independent of the initial interest rate curve.

This holds true for a number of interesting model choices (for example Gaussian models and some stochastic volatility models) but is definitely restrictive. This assumption can be relaxed dramatically; this is left for future sections to explore.

In this light, consider a perturbed interest rate model (compare to (3.1))

$$\begin{aligned} dP^\varepsilon(t,T) &= r^\varepsilon(t) P^\varepsilon(t,T)\,dt + P^\varepsilon(t,T)\,\sigma(t,T)\,dW(t), \\ P^\varepsilon(0,T) &= P_0^\varepsilon(T), \quad T \geq 0. \end{aligned} \tag{10.1}$$

Let us decorate a price process of any contingent claim obtained in the model (10.1) with a superscript ε. In particular let $H_n^\varepsilon(t)$ be the value, at time t, of the Bermuda swaption H_n in the perturbed model (10.1).

For a price process of any contingent claim $\xi(t)$, we define its market delta with respect to the shock θ by

$$\partial_\theta \xi(t) = \lim_{\varepsilon \downarrow 0} \varepsilon^{-1} \left(\xi^\varepsilon(t) - \xi^0(t) \right).$$

We call this quantity a "market" delta because this is the delta, or sensitivity, that is required in practice to manage interest rate risk. It is a sensitivity to a shock of an initial interest rate curve, as opposed to sensitivities to stochastic shocks of random variables, something we called "model" ("underlying" and "discount") deltas before.

The following theorem applies the technology developed so far to derive an expression for market deltas of the Bermuda swaption.

Theorem 10.1. *The market delta of the Bermuda swaption H_0 to an initial interest rate curve shock θ, under the condition that the volatility process is independent of the initial* *interest rate curve, can be computed as follows,*

$$\begin{aligned}(10.2)\quad \partial_\theta H_0(0) &= \sum_{n=1}^{N-1}\int_X 1_{\{R_n\}}(x)\times\partial_\theta E_n(x,T_n)\times\ \psi(x;T_n)\ \nu(dx)\\ &\quad-\sum_{n=0}^{N-2}\tau_{n+1}\int_X 1_{\{R_n^c\}}(x)\times\partial_\theta F(x,T_n,T_{n+1})\times\ \psi(x;T_n)\ \nu(dx).\end{aligned}$$

Here $\psi(x;t)$ is the survival density, R_n is the "exercise" region $\{x\in X: H_n(x,T_n)\le E_n(x,T_n)\}$ at time T_n in the state space X, $\partial_\theta E_n(x,T_n)$ is the market delta of the exercise value expressed as a function of the state process $x(T_n)$, and $\partial_\theta F(x,T_n,T_{n+1})$ is the market delta of the future discount rate expressed as a function of the state process $x(T_n)$.

Proof. The theorem follows from the chain rule by using $\partial_\theta E_n(T_n)$, $\partial_\theta F(T_n,T_{n+1})$ as shocks d_n^u, d_n^d in Corollary 8.3. We use the fact that $\sigma(t,T)$ is independent of ε when applying the chain rule. ∎

This theorem is the most practically important result of this paper. It shows that, once the survival density is determined, all market deltas (i.e. market deltas for any shock θ) can be computed by simple integration of the survival density against market deltas of the exercise values and discount rates. The market deltas of exercise values are, as a rule, computationally inexpensive to compute. Recall that E_n's are linear combinations of zero coupon bonds, so all we really need to be able to compute are $\partial_\theta P(t,T)$. If a closed-form formula is available for $P(t,T)$, it can usually be differentiated with respect to θ to obtain a closed-form formula for $\partial_\theta P(t,T)$. Later in the paper we present a case study for one of the models, where we work out all the details of these calculations.

Remark 10.1. *A word of caution is in order. In the recursion*

$$H_n(T_n)=e^{-\tau_{n+1}F(T_n,T_{n+1})}\mathbf{E}_{T_n}^{T_{n+1}}\max\{H_{n+1}(T_{n+1}),E_{n+1}(T_{n+1})\},$$

there are three "components" that are affected by the shock θ. It is the exercise value $E_{n+1}(T_{n+1})$, the discount rate $F(T_n,T_{n+1})$ and the forward measure $\mathbf{P}_{T_n}^{T_{n+1}}$. The contributions of the first two are accounted for in Theorem 10.1. We also assumed that $\sigma(t,T)$ is unaffected by the shock θ. This means that the risk-neutral measure $\mathbf{P}_{T_n}$ is unaffected by the shock θ. However, it ***does not guarantee*** *that the forward measure $\mathbf{P}_{T_n}^{T_{n+1}}$ is unaffected. The differences between the two measures are so small from the point of view of delta computations that we ignore it.*

10.2. **Multiple shocks.** Theorem 10.1 derives the algorithm for computing any market delta. The algorithm is very efficient in the situations where more than one market delta is required. Computational savings come from having to compute the survival density only once.

In certain situations, computational savings can be pushed even further. Extra computational savings come from avoiding performing all integrations in (10.2).

Various schemes are possible. Usually the best scheme can be designed only after carefully analyzing a structure of a particular model. The following approach, however, should provide a good starting point for exploration.

Computing a single "model" delta requires only one integration. Computing a single "market" deltas requires $2(N-1)$ integrations, since each shock θ potentially affects all underlying values E_n, $n = 1, \ldots, N-1$, and all discount rates F_n, $n = 0, \ldots, N-2$. However, if we choose the shock θ carefully so it affects only a few of E_n and F_n (or at least affects only a few of them *significantly*), then we can cut down on the number of integrations performed.

Shocks $\theta_i(\cdot)$, $1 \leq i \leq I$, are usually fixed of course. These are the shocks that traders would like to see their risk expressed in. However, we can always define a new set,

$$\theta_i'(\cdot), 1 \leq i \leq I,$$

by

$$\theta_i'(\cdot) = \theta_j'(\theta_1(\cdot), \ldots, \theta_I(\cdot)), \quad 1 \leq i \leq I.$$

As long as the Jacobian $\frac{\partial \theta_i'}{\partial \theta_j'}$ is invertible, the deltas to the shocks $\theta_i'(\cdot)$ can be computed instead, and then converted to the deltas to the shocks $\theta_i(\cdot)$ by simple matrix manipulations. The choice of the "intermediate" shocks $\theta_i'(\cdot)$ can be made to minimize the number of integrations needed to compute the deltas to them.

As we mentioned in the beginning of this section, a particular choice of such shocks will depend on the structure of the model. Here we present one as an example. We believe it should work reasonably well for a good selection of models.

We use shocks to the initial interest rate curve parametrized as zero-coupon bonds:

$$P_0^{\varepsilon}(T) = P_0(T) + \varepsilon\theta(T), \quad T \geq 0.$$

We choose i, $1 \leq i \leq N-1$, and set the shock $\theta_i(T_k)$ according to the following rule (here c is the fixed rate of all underlying swaps, see (4.1)):

$$\begin{aligned}
\theta_i(T_k) &= 0, \quad i < k \leq N, \\
\theta_i(T_i) &= 1, \\
\theta_i(T_k) &= -c \sum_{j=k+1}^{n} \theta_i(T_j), \quad 1 \leq k < i.
\end{aligned}$$

We mention in passing that

$$\theta_i(T_k) = 1 - c(1-c)^{i-k-1}, \quad 1 \leq k < i.$$

We use ∂_{θ_k} to denote the derivative in the direction of the k-th shock θ_k.

It is easy to see that for all k, i, for underlying swaps observed at time T_0,

$$\partial_{\theta_i} E_n(T_0) = 1_{\{i=n\}}.$$

Henceforth, it is reasonable to expect that a similar equality will (approximately) hold for the underlying swap values on the exercise dates,

$$\partial_{\theta_i} E_n(T_n) \approx 1_{\{i=n\}}.$$

At least one can expect that

$$\partial_{\theta_i} E_n(T_n) \ll \partial_{\theta_i} E_i(T_i).$$

Therefore, in the formula

$$\begin{aligned}(10.3)\quad \partial_{\theta i} H_0(0) &= \sum_{n=1}^{N-1} \int_X 1_{\{R_n\}}(x) \times \partial_{\theta_i} E_n(x, T_n) \times \psi(x; T_n)\, \nu(dx) \\ &\quad - \sum_{n=0}^{N-2} \tau_{n+1} \int_X 1_{\{R_n^c\}}(x) \times \partial_{\theta_i} F(x, T_n, T_{n+1}) \times \psi(x; T_n)\, \nu(dx).\end{aligned}$$

the first sum will be dominated by the term corresponding to $n = i$,

$$\int_X 1_{\{R_i\}}(x) \times \partial_{\theta_i} E_i(x, T_i) \times \psi(x; T_i)\, \nu(dx).$$

This term can be computed first. Assuming that the further n is from i, the smaller the contribution of the n-th term is to the i-th market deltas, an adaptive scheme in which the terms are computed as long as their contribution does not fall below a certain threshold can be implemented.

The impact of the proposed shocks θ_i on discount rates can be assessed in a similar way. Recall that the forward discount rates observed at T_0 are give by

$$F(T_0, T_n, T_{n+1}) = -\frac{1}{\tau_{n+1}} \log P(T_0, T_n, T_{n+1}),$$

so that

$$\begin{aligned}\partial_{\theta_i} F(T_0, T_n, T_{n+1}) &= -\frac{1}{\tau_{n+1} P(T_0, T_n, T_{n+1})} \partial_{\theta_i} P(T_0, T_n, T_{n+1}) \\ &= -\frac{1}{\tau_{n+1}} \left(\frac{\theta_i(T_{n+1})}{P(T_0, T_{n+1})} - \frac{\theta_i(T_n)}{P(T_0, T_n)} \right).\end{aligned}$$

Of all $\partial_{\theta_i} F(T_0, T_n, T_{n+1})$, the two biggest are $i = n$ and $i = n + 1$. Hence, it is reasonable to expect that the biggest contributions to the second sum in (10.3) are likely to be from the terms that correspond to $n = i$ and $n + 1 = i$. Again assuming monotonicity between $\partial_{\theta_i} F(x, T_n, T_{n+1}) = \partial_{\theta_i} F(x, T_n, T_n, T_{n+1})$ for various n, an adaptive scheme with term cutoff can be implemented.

11. Gammas

A gamma is a sensitivity of a delta with respect to a shock to the initial interest rate curve. If there are I shocks $\theta_i(\cdot)$, $1 \le i \le I$, there are I^2 gammas, $\partial_{\theta_j} \partial_{\theta_i} H_0(0)$. It is not uncommon for the gammas to be aggregated in some way. A trader may request a single gamma number which is the sum of all gammas,

$$\sum_{i,j=1}^{I} \partial_{\theta_j} \partial_{\theta_i} H_0(0).$$

Another popular way of aggregating gammas is "by row" of the gamma matrix,

$$(11.1)\qquad \sum_{i=1}^{I} \partial_{\theta_j} \partial_{\theta_i} H_0(0).$$

This has a meaning of "sensitivity of each bucketed delta with respect to a combined shock $\sum_{i=1}^{I}\theta_i(\cdot)$". Since the gamma matrix can be quite big, such aggregation reduces information overload. It also reduces numerical noise inherent in gamma calculations.

If aggregated gammas are requested, such as in (11.1), then we advocate using the following procedure for their computation. We apply a shock $\varepsilon\sum_{i=1}^{I}\theta_i(\cdot)$ for some small $\varepsilon>0$ to the initial curve and compute all bucketed deltas as in the previous section. Then we apply a shock $-\varepsilon\sum_{i=1}^{I}\theta_i(\cdot)$ to the initial interest rate curve and compute bucketed deltas as in the previous section. Then we subtract "down" deltas from "up" deltas, divide by 2ε and obtain an estimate of all bucketed aggregated gammas.

If the whole gamma matrix is required, we have two choices. The first approach is to do the same thing as for aggregated bucketed gammas above, only applying each shock θ_i individually to the initial term curve. The second approach is to try to derive recursive relations for gammas the same way we did it for deltas, and try to explore that representation.

It will become clear shortly that the second approach does not offer much advantage in terms of computational effort (it can however be more accurate). We sketch the second approach here without filling in all the details, as they are basically the same as for the algorithm for computing deltas.

Consider the recursive formula for $\Delta_n^u h_k$ obtained in Theorem 7.2 for $k<n-1$,

$$\Delta_n^u h_k(A)=\pi_{T_k}\left(1_{\{H_{k+1}(T_{k+1})>E_{k+1}(T_{k+1})\}}\Delta_n^u h_{k+1}(A)\right).$$

Applying Δ_m^u to this relation formally for $m<n-1$, we get

$$\begin{aligned}\Delta_m^u\Delta_n^u h_k(A) &= -\pi_{T_k}\left(\delta_{\{H_{k+1}(T_{k+1})=E_{k+1}(T_{k+1})\}}\Delta_m^u h_{k+1}(A)\,\Delta_n^u h_{k+1}(A)\right)\\ &\quad+\pi_{T_k}\left(1_{\{H_{k+1}(T_{k+1})>E_{k+1}(T_{k+1})\}}\Delta_m^u\Delta_n^u h_{k+1}(A)\right).\end{aligned}\tag{11.2}$$

The quantity

$$\delta_{\{H_{k+1}(T_{k+1})=E_{k+1}(T_{k+1})\}}$$

in (11.2) is a delta function for the event $H_{k+1}(T_{k+1})=E_{k+1}(T_{k+1})$. Under the appropriate regularity conditions on $H_{k+1}(T_{k+1})$, $E_{k+1}(T_{k+1})$ (basically their difference needs to have a density) the expression

$$\pi_{T_k}\left(\delta_{\{H_{k+1}(T_{k+1})=E_{k+1}(T_{k+1})\}}\times\Delta_m^u h_{k+1}(A)\times\Delta_n^u h_{k+1}(A)\right)$$

can be defined rigorously. Without going into much theoretical detail, consider the case when the underlying process $x(t)$ is one-dimensional, $\phi_{s,x}(y;t)$ is as defined in Section 9, and, for each k, there is only one point x_k^* such that $H_k(x_k^*,T_k)=E_k(x_k^*,T_k)$. Then

$$\begin{aligned}&\pi_{T_k}\left(\delta_{\{H_{k+1}(T_{k+1})=E_{k+1}(T_{k+1})\}}\times\Delta_m^u h_{k+1}(A)\times\Delta_n^u h_{k+1}(A)\right)\\ &\qquad=\phi_{T_k,x(T_k)}\left(x_{k+1}^*;t\right)\times\Delta_m^u h_{k+1}\left(x_{k+1}^*\right)\times\Delta_n^u h_{k+1}\left(x_{k+1}^*\right).\end{aligned}$$

This relation, and similar ones for Δ_m^d, can be iterated to obtain recursive formulas for the Gammas of $H_0(T_0)$ with respect to shocks d_n^u, d_n^d. They are presented in Appendix B. One of the recursive formulas reads, for $m\le n$,

(11.3)

$$\Delta_m^u \Delta_n^u H_0(T_0) = -\sum_{k=1}^{m-1} \pi_0 \left(\left(\prod_{i=1}^{k-1} 1_{\{H_i(T_i)>E_i(T_i)\}} \right) \delta_{\{H_k(T_k)=E_k(T_k)\}} \Delta_m^u H_k(T_k)\, \Delta_n^u H_k(T_k) \right) + \pi_0 \left(\left(\prod_{i=1}^{m-1} 1_{\{H_i(T_i)>E_i(T_i)\}} \right) \delta_{\{H_m(T_m)=E_m(T_m)\}} d_m^u \left(1_{\{m=n\}} d_m^u - 1_{\{m\neq n\}} \Delta_n^u H_m(T_m) \right) \right).$$

In the one-factor setting mentioned above, the formula takes the form

$$\begin{aligned} \Delta_m^u \Delta_n^u h_0(A) &= -\sum_{k=1}^{m} \psi(x_k^*; T_k)\; \Delta_m^u H_k(x_k^*, T_k)\; \Delta_m^u H_k(x_k^*, T_k) \\ &\quad + \psi(x_m^*; T_m)\, d_m^u(x_m^*) \left(1_{\{m=n\}} d_m^u(x_m^*) - 1_{\{m\neq n\}} \Delta_n^u H_m(x_m^*, T_m) \right). \end{aligned}$$

In this formula the deltas of hold values $\Delta_m^u H_k$, $\Delta_m^u H_k$ evaluated at the exercise boundary point x_k^* are needed. These values can be computed in the same way as $\Delta_m^u H_0(T_0)$, by integrating known payoffs against a survival density *started at* (x_k^*, T_k). Let us briefly explain what we mean.

For any (x, s) an optimal exercise time index starting at time s, is defined by

$$\eta(s) = \min\{n \geq 1 : H_n(T_n) \leq E_n(T_n),\, T_n > s\}.$$

A survival measure started at (x, s), is defined by

$$\Psi_{x,s}(Y;t) = \mathbf{E}_{s,x}\left(B_t^{-1} \times 1_{\{\eta(s)>n(t)\}} \times 1_{\{x(t)\in Y\}}\right)$$

for

$$Y \in \mathcal{B}(X),$$

and its density, a survival density started at (x, s), by

$$\Psi_{x,s}(Y;t) = \int_Y \psi_{x,s}(y;t)\; \nu(dy).$$

It is clear that, to compute the whole gamma matrix, one needs to compute survival densities that start at $(x(T_0), T_0)$ (*the* survival density used for computing deltas) plus $N-2$ survival densities starting at (T_k, x_k^*), $k = 1, \ldots, N-2$. To obtain each survival density $\psi_{x_k^*,T_k}(y;t)$, a PDE from Theorem 9.2 with a different initial condition $\psi_{x_k^*,T_k}(y;T_k) = \delta_{x_k^*}(y)$ needs to be solved.

Even in this simple situation of a one-factor state process $x(\cdot)$, there does not appear to be any computational advantage in using our approach over a simple scheme where the gammas are computed by numerical differentiation of deltas that are computed as in Section 10.

12. Vegas

This section discussed the extension of our methodology to the problem of computing vegas, sensitivities to changes in interest rates volatilities.

12.1. **Conditions on the volatility process.** To be able to say something meaningful about computing vegas,the universe of all possible volatility processes $\sigma(t,T)$ (see (3.1)) needs to be restricted. The volatility process is required to satisfies two conditions, that of *completeness* and *locality*, as defined below.

It is market convention to quote prices of European swaptions in terms of Black volatilities, i.e. volatilities implied from swaption prices using the Black model. For that reason, it is quite common to compute volatility sensitivities with respect to these volatilities. Let us denote a set of Black volatilities to which the vegas are required by $\mathcal{V}$. Let us denote the j-th Black volatility in this set by u_j^*, $1 \le j \le J$.

Typically, the volatility process $\sigma(t,T)$ has a number of adjustable parameters. For the purposes of Bermuda swaption valuation, they are usually chosen such that the model prices for all European swaptions in the set $\mathcal{V}$ coincide with their market prices. This process is known as volatility calibration.

As we already mentioned, the vegas that are usually required are sensitivities to the shocks of market (Black) volatilities of European swaptions (we will call these "market" vegas). From the model prospective, however, the sensitivities that are easier to obtain are vegas with respect to parameters of the volatility process $\sigma(t,T)$ (we will call these "model" vegas).

We call a volatility process $\sigma(t,T)$ *complete* if there is enough flexibility in the model's volatility parameters to represent independent shocks to the market volatilities of the required subset of European swaptions.

Let $v = \{v_1, \ldots, v_I\}$ be the parameters of $\sigma(t,T)$, i.e.

$$\sigma(t,T) = \sigma(\omega, t, T, v).$$

Let $S \subset \mathbb{R}^I$ be a subset of volatility parameter values $\{v_1, \ldots, v_I\}$ for which the volatility process $\sigma(\omega, t, T, v)$ is well defined and satisfies the necessary regularity conditions. We assume that S is an open set. Let us define a mapping from the model's volatility parameters to Black volatilities for the set $\mathcal{V}$ of European swaptions:

$$u : S \to \mathbb{R}^J,$$

where by definition $u_j(v)$ is the Black volatility of the j-th European swaption in the set $\mathcal{V}$ implied from the model's price computed with volatility parameters v.

We denote by v^* a point in S such that

$$u(v^*) = u^*.$$

We assume that $u(v)$ is differentiable at $v = v^*$.

Definition 12.1. *The volatility process $\sigma(\omega, t, T, v)$ (and, correspondingly, the model) is* complete *if u^* is an interior point of the image of $u(S)$ of S under the mapping u.*

We define $\nabla_i \xi(t)$ to be the sensitivity of a price of a contingent claim with a price process $\xi(t)$ to a shock to volatility parameter v_i.

Definition 12.2. *We call a volatility process $\sigma(\omega, t, T, v)$* local *if for each i, $1 \le i \le I$, there exists $n = n(i)$, $1 \le n(i) \le N-1$, such that for any $\mathcal{F}_T$-measurable random variable ξ,*

$$\nabla_i \pi_t(\xi) = 0 \text{ for } t \ge T_{n(i)+1} \text{ or } T \le T_{n(i)}. \tag{12.1}$$

What this definition means is that we require a shock to the parameter v_i be confined, or local, it its effect to the interval $[T_n, T_{n+1}]$ for some n.

The conditions of completeness and locality are quite general. There is a large number of interesting interest rate models that satisfy them. A realistic example of such a model is given in Section 15.

The locality condition is necessary for successful application of our method to computing model vegas. The completeness condition is needed to be able to translate model vegas into market vegas.

12.2. **Computationally efficient formulas for model vegas.** The following theorem is a counterpart of Theorem 7.2 for model vegas.

Theorem 12.1. *Let i be such that $1 \leq i \leq I$. Let $n = n(i)$ be as defined in (12.1) in Definition 12.2. Let k be such that $0 \leq k \leq N-1$. Then*

1. For $k > n(i)$,

$$\nabla_i H_k(T_k) = 0;$$

2. For $k < n(i)$,

$$\begin{aligned} \nabla_i H_k(T_k) &= \pi_{T_k}\left(1_{\{H_{k+1}(T_{k+1}) > E_{k+1}(T_{k+1})\}} \times \nabla_i H_{k+1}(T_{k+1})\right) \\ &\quad + \pi_{T_k}\left(1_{\{H_{k+1}(T_{k+1}) \leq E_{k+1}(T_{k+1})\}} \times \nabla_i E_{k+1}(T_{k+1})\right). \end{aligned} \tag{12.2}$$

The second term in (12.2) appears because, technically, the exercise value $E_{k+1}(T_{k+1})$ can depend on volatility. This dependence, however, is typically very mild, because there is very little "curvature" in the underlying swap payoffs. The extra contribution of this term to vega can usually be safety ignored. With this in mind, we get the following corollary to Theorem 12.1, a counterpart of Corollary 8.3 for deltas.

Corollary 12.2. *Ignoring the contributions of exercise value vegas to Bermuda swaption model vegas and setting $n = n(i)$ as defined in Definition 12.2,*

$$\nabla_i H_0(T_0) = \pi_0\left(1_{\{H_1(T_1) > E_1(T_1)\}} \times \cdots \times 1_{\{H_n(T_n) > E_n(T_n)\}} \times \nabla_n H_n(T_n)\right).$$

In particular, this expression can be computed by integrating $\nabla_i H_n(T_n)$ against the survival density,

$$\nabla_i H_0(T_0) = \int_X 1_{\{R_n^c\}}(x) \times \nabla_i H_n(x, T_n) \times \psi(x; T_n)\, \nu(dx).$$

The difference between Theorem 12.1 for vegas and Theorems 7.2, 7.3 is that a closed-form formula for $\nabla_i H_{n(i)}(T_n)$ (the counterpart of the case $k = n$ in Theorem 12.1) is not available. It has to be computed numerically by rolling back the payoff

$$\max\left(H_{n(i)+1}\left(T_{n(i)+1}\right), E_{n(i)+1}\left(T_{n(i)+1}\right)\right)$$

from time $T_{n(i)+1}$ to $T_{n(i)}$ with a shocked value of the volatility parameter v_i. Note that this is only a "one-step" rollback, from $T_{n(i)+1}$ to $T_{n(i)}$, as opposed to the full rollback from $T_{n(i)}$ to $T_0 = 0$. In a typical PDE implementation, the numerical effort required is linear in the number of time nodes, so this scheme realizes a speed-up of the order N.

12.3. **Mapping model vegas to market vegas.** The function $u(v)$ maps model volatility parameters to Black volatilities of European swaptions. The right-inverse of $u(\cdot)$, denoted by $v(u)$ and defined by

$$u(v(x)) = x \quad \forall x \in u(S),$$

defines a process of calibration. For a given collection of market (Black) volatilities u, one can use $\sigma(t, T, v(u))$ in Bermuda swaption valuation. This provides a mapping between market Black volatilities and the model's Bermuda swaption value

$$u \longmapsto H_0(T_0).$$

We assume this function is differentiable. By the chain rule

$$\begin{aligned} \nabla_i H_0(T_0) &= \frac{\partial}{\partial v_i} H_0(T_0) \\ &= \sum_{j=1}^{J} \frac{\partial}{\partial u_j} H_0(T_0) \frac{\partial u_j}{\partial v_i}, \end{aligned} \tag{12.3}$$

for $1 \leq i \leq I$.

The left-hand sides of the equations (12.3) are known from the previous section. The matrix $\left\{\frac{\partial u_j}{\partial v_i}\right\}$ can be computed by bumping volatility parameters v, revaluing European swaptions in the model, and computing $u(v)$ by implying Black volatilities from them. The linear equations (12.3) can be solved (here we use the completeness condition) to get

$$\left.\frac{\partial}{\partial u_j} H_0(T_0)\right|_{u=u^*},$$

the market vegas.

The only extra complication we need to be aware of is that I can be larger than J. In this case, the equations (12.3) should be understood in the least-squares sense.

13. Deltas for general volatility processes

In Section 10 the volatility process $\sigma(t, T)$ was restricted to not have any dependence on the interest rate curve

This assumption is relaxed significantly in this section. Instead of complete independence a less restrictive assumption of local dependence on the rates is made, very much in line with the assumptions we imposed in Section 12.

Let $v(\omega, t, P(\cdot, \cdot))$ be an $\mathbb{R}^I$-valued $\mathcal{H}_t$-adapted stochastic process that depends on the trajectories of the zero coupon bond processes (potentially for all maturities).

We assume that $\sigma(t, T)$ is of the form

$$\sigma = \sigma(\omega, t, T, v(\omega, t, P(\cdot, \cdot)))$$

where $\sigma(\omega, t, T, v)$ for each fixed $v \in \mathbb{R}^I$ is a local volatility process as defined in Definition 12.2.

We encapsulate interest rate dependence of volatility in parameters in which the volatility process is local.

Such a definition covers a wide range of interesting and useful interest rate models. See an example in Section 15.

Recall from Section 10 that we used the notation ∂_θ for the sensitivity with respect to a θ-shock of the initial interest rate curve.

A full derivative of H_0 with respect to the θ-shock can be computed as a sum of two parts. The first part is the delta, as obtained in Section 10, with the volatility process $\sigma(\omega, t, T, v(\omega, t, P(\cdot,\cdot)))$ kept constant. The second part is the sensitivity to volatility process changes that result from its dependence on the θ-shock, while keeping the exercise values and discount rates constant. The first part is computed in Theorem 10.1. The second part is related to vegas. Since we imposed a locality condition on the volatility process, this contribution can be computed as in Corollary 12.2.

Let us formalize these considerations. Recall that we used the notation $P^\varepsilon(\cdot,\cdot)$ for the values of zero coupon discount bonds in the model with the shocked initial condition $P_0(\cdot) + \varepsilon\theta(\cdot)$ and with the volatility process σ not shocked. Define

(13.1)
$$\nabla_n^\theta H_n(T_n) = \lim_{\varepsilon \downarrow 0} \varepsilon^{-1} \left(\pi_{T_n}^\varepsilon \max\left(H_{n+1}(T_{n+1}), E_{n+1}(T_{n+1})\right) - \pi_{T_n} \max\left(H_{n+1}(T_{n+1}), E_{n+1}(T_{n+1})\right)\right),$$

where $\pi_{T_n}^\varepsilon$ is computed with $\sigma = \sigma(\omega, t, T, v(\omega, t, P^\varepsilon(\cdot,\cdot)))$.

Theorem 13.1. *The market delta of the Bermuda swaption H_0 to an initial interest rate curve shock θ, is computed as follows,*

$$\begin{aligned}
\partial_\theta H_0(0) &= \sum_{n=1}^{N-1} \int_X 1_{\{R_n\}}(x) \times \partial_\theta E_n(x, T_n) \times \psi(x; T_n)\, \nu(dx) \\
&\quad - \sum_{n=0}^{N-2} \tau_{n+1} \int_X 1_{\{R_n^c\}}(x) \times \partial_\theta F(x, T_n, T_{n+1}) \times H_n(x, T_n) \times \psi(x; T_n)\, \nu(dx) \\
&\quad + \sum_{n=1}^{N-1} \int_X 1_{\{R_n^c\}}(x) \times \nabla_n^\theta H_n(x, T_n) \times \psi(x; T_n)\, \nu(dx).
\end{aligned}$$

Here it is assumed, just like in Corollary 12.2, that the contribution of sensitivities of exercise values with respect to changes in volatility can be ignored

Note that the quantities $\nabla_n^\theta H_n(x, T_n)$ are computed numerically via (13.1) in a one-step rollback from T_{n+1} to T_n.

14. Deltas and volatility calibration

Recall from Section 12 that we denoted the model's volatility parameters by $v = \{v_1, \ldots, v_I\}$, and the market's Black volatilities of European swaptions to which the model is calibrated by u_j^*, $1 \le j \le J$. Also recall the mapping from the model's volatility parameters to Black volatilities for the set of European swaptions:

$$u : S \to \mathbb{R}^J,$$

where $u_j(v)$ is the Black volatility of the j-th European swaption implied from the model's price computed with volatility parameters v use, and the definition of v^* by

$$u(v^*) = u^*.$$

It is quite common for the mapping $u(\cdot)$ to depend on the interest rate curve.

The "market" deltas for which the algorithm was derived in previous sections are sensitivities of the price of a Bermuda swaption to interest rate shocks assuming that *the model's volatility parameters* v are constant. What sometimes is required is a sensitivity of as Bermuda swaption price to a shock θ assuming the *market volatilities are constant.* In the case when $u(\cdot)$ depends on interest rates, the two will not be the same. It is, however, quite easy to translate one into another. Intuitively, the latter is a combination of the former and the vegas.

We present the translation in symbolic form. Let r denote the initial interest rate curve, v and u as above, $f(r,v)$ the value of the Bermuda swaption with rates r and volatilities v. Let us specify the dependence of the functions $u(\cdot)$ and $v(\cdot)$ on the rates explicitly,

$$\begin{aligned} v &= v(r,u), \\ u &= u(r,v), \\ u(r,v(x)) &= x. \end{aligned} \tag{14.1}$$

Then the total derivative of f with respect to r is equal to the sum of partial derivatives,

$$\frac{d}{dr} f(r,v(r,u)) = \frac{\partial}{\partial r} f(r,v) + \frac{\partial}{\partial v} f(r,v) \times \frac{\partial}{\partial r} v(r,u). \tag{14.2}$$

Furthermore, differentiating (14.1),

$$\frac{d}{dr} u(r,v(x)) = \frac{d}{dr} x = 0,$$

which leads to

$$\frac{\partial}{\partial r} u(r,v(x)) + \frac{\partial}{\partial v} u(r,v(x)) \times \frac{\partial}{\partial r} v(x) = 0.$$

Thus

$$\frac{\partial}{\partial r} v(u) = -\left[\frac{\partial}{\partial v} u(r,v)\right]^{-1} \left[\frac{\partial}{\partial r} u(r,v)\right].$$

Plugging the last formula into (14.2) and evaluating it at $u = u^*$ we obtain,

$$\begin{aligned} \left.\frac{d}{dr} f(r,v(r,u))\right|_{u=u^*} &= \frac{\partial}{\partial r} f(r,v^*) - \frac{\partial}{\partial v} f(r,v^*) \times \mu, \\ \mu &= \left[\frac{\partial}{\partial v} u(r,v)\right]^{-1} \left[\frac{\partial}{\partial r} u(r,v)\right]. \end{aligned} \tag{14.3}$$

This formula reads: the delta while keeping market volatilities u constant is equal to the delta with model volatilities v constant minus the model vega times a coefficient μ. The coefficient μ is a ratio of the derivatives of market volatilities with respect to rates and to model volatilities (in a typically multidimensional setting it is a product of the matrix inverse $\left[\frac{\partial}{\partial v} u(r,v)\right]^{-1}$ and a vector $\left[\frac{\partial}{\partial r} u(r,v)\right]$). Volatility calibration of a model usually relies on having a fast and efficient formula for $u(r,v)$. Therefore the coefficient μ can usually be obtained at a negligeable computational cost.

15. Case study: The Cheyette Model

15.1. **The Cheyette model definition.** In this section we look at a specific model and show how our method can be applied in practice. The method is demonstrate for the so-called Cheyette model (see [Che91] and [And01b]). It is chosen because it is a viable model for valuing Bermuda swaptions (see [And01a]) and relatively easy to describe.

The simplest model we could have applied our method to, the Hull-White model, is a special case of the Cheyette model. The Cheyette model adds the dependence of volatility on the interest rates thus making the example more interesting.

Let us quickly recap the definition of the Cheyette model. The Cheyette model is most conveniently written in terms of two state variables, $x(t)$ and $y(t)$, that follow the dynamics

$$\begin{aligned} dx(t) &= (y(t) - \kappa(t)\, x(t))\; dt + \eta(t, x(t), y(t))\; dW(t), \\ dy(t) &= \left(g(t)^2 h(t, x(t), y(t))^2 - 2\kappa(t)\, y(t)\right)\; dt. \end{aligned}$$

The bond reconstruction formulas

$$P(t,T) = P(x(t), y(t), t, T)$$

are given by

$$P(x,y,t,T) = \frac{P(0,T)}{P(0,t)} \exp\left(-G(t,T)\, x - \frac{1}{2} G(t,T)^2\, y\right), \tag{15.1}$$

where

$$\begin{aligned} G(t,T) &= \int_t^T \frac{g(s)}{g(t)}\, ds, \\ g(s) &= \exp\left(-\int_0^s \kappa(u)\; du\right). \end{aligned}$$

The function $\kappa(u)$ is interpreted as a (time-dependent) mean reversion. In terms of the zero coupon bonds, the model can be written as

$$dP(t,T) = r(t)\, P(t,T)\; dt + \sigma(t,T)\, P(t,T)\; dW(t),$$

where

$$r(t) = f(0,t) + x(t),$$

and

$$\sigma(t,T) = -G(t,T)\, \eta(t, x(t), y(t)).$$

A particular specification of the volatility function η is chosen,

$$\begin{aligned} \eta(t, x(t), y(t)) &= \lambda(t)\, \gamma(t, x(t), y(t)), \\ \lambda(t) &= v_k, \quad t \in (T_k, T_{k+1}], \\ \gamma(t, x(t), y(t)) &= F_{k+1}^{\alpha}(t), \quad t \in (T_k, T_{k+1}], \\ F_k(t) &= \frac{P(t,T_k) - P(t,T_N)}{(T_N - T_k)\, P(t,T_N)}, \\ \alpha &\in [0,1]. \end{aligned} \tag{15.2}$$

Here a CEV-type specification with parameter $\alpha.is$ used, with a twist. The twist is in using different rates to apply the power α to. A simple rate covering the period $[T_k, T_N]$ is used for t in the time interval $(T_{k-1}, T_k]$. This specification is known to produce volatility skews that does not "die out" as time to expiry increases, a common problem with standard CEV specifications.

Numbers v_k are instantaneous volatilities that are chosen to fit market prices (or Black volatilities) of "core" swaptions, i.e. swaptions with expiries T_k covering periods $[T_k, T_N]$ for all k, $1 \leq k \leq N-1$. We denote their Black volatilities by u_k^*.

Any contingent claim price $V(t, x, y)$, as a function of the state variables $x(t)$ and $y(t)$, satisfies the equation

$$\frac{\partial}{\partial t} V(t, x, y) + (\Lambda V)(t, x, y) = (f(0, t) + x) V(t, x, y), \tag{15.3}$$

where the infinitesimal generator of the Markov process $(x(\cdot), y(\cdot))$ is given by

$$\begin{aligned}(\Lambda V)(t, x, y) &= (y - \kappa(t) x) \frac{\partial}{\partial x} V(t, x, y) \\ &\quad + \left(g^2(t) h^2(t, x, y) - 2\kappa(t) y\right) \frac{\partial}{\partial y} V(t, x, y) \\ &\quad + \frac{1}{2} h^2(t, x, y) \frac{\partial^2}{\partial x^2} V(t, x, y).\end{aligned} \tag{15.4}$$

15.2. **Equation for the survival density.** The equation for the survival density follows from the stochastic differential equations for the Markov state variables $x(t)$ and $y(t)$ presented above. If $\psi(t, x, y)$ is the survival density, we have for $t \in (T_n, T_{n+1}]$ for each n,

$$\begin{aligned}\frac{\partial}{\partial t} \psi(t, x, y) &= -\frac{\partial}{\partial x} \left[(y - \kappa(t) x) \psi(t, x, y)\right] \\ &\quad - \frac{\partial}{\partial y} \left[\left(g^2(t) h^2(t, x, y) - 2\kappa(t) y\right) \psi(t, x, y)\right] \\ &\quad + \frac{1}{2} \frac{\partial^2}{\partial x^2} \left[h^2(t, x, y) \psi(t, x, y)\right] - (f(0, t) + x) \psi(t, x, y).\end{aligned}$$

On the dates T_n, $1 \leq n \leq N-1$, the conditions are specified in Theorem 9.2 (these are the initial conditions for the PDE above when solved for the interval $(T_n, T_{n+1}]$):

$$\psi(T_n + 0, x, y) = \psi(T_n, x, y) \times 1_{\{(x,y) \in R_n^c\}},$$

where, for each n, R_n is the "exercise" region and R_n^c is the "hold" region:

$$\begin{aligned}R_n &= \left\{(x, y) \in \mathbb{R}^2 : H_n(x, y, T_n) \leq E_n(x, y, T_n)\right\}, \\ R_n^c &= \left\{(x, y) \in \mathbf{R}^2 : H_n(x, y, T_n) > E_n(x, y, T_n)\right\}.\end{aligned}$$

Here $E_n(x, y, T_n)$ is the n-th exercise value at time T_n at the node $x(T_n) = x$, $y(T_n) = y$, and $H_n(x, y, T_n)$ is the n-th hold value at time T_n at the node $x(T_n) = x$, $y(T_n) = y$.

At time $t = T_0 = 0$, we have

$$\psi(t, x, y) = \delta_{(0,0)}(x, y).$$

15.3. **Deltas to initial interest rate curve shocks.** Without loss of generality, we use shocks suggested in Section 10.2 (as explained in that section, deltas to other shocks are computed by matrix multiplication of deltas to these shocks). We choose k, $1 \leq k \leq N-1$, and set the shock $\theta_k(T_n)$ according to the following rule (here c is the fixed rate on the swaps, see (4.1)):

$$\begin{aligned}
\theta_k(T_n) &= 0, \quad k < n \leq N, \\
\theta_k(T_k) &= 1, \\
\theta_k(T_n) &= -c \sum_{i=n+1}^{k} \theta_k(T_i), \quad 1 \leq n < k.
\end{aligned}$$

We use ∂_{θ_k} to denote the derivative in the direction of the k-th shock θ_k.

Recall from Section 10.2 that these shocks satisfy

$$\partial_{\theta_k} E_n(T_0) = 1_{\{k=n\}} \tag{15.5}$$

for all k, n.

Let us compute the effect of these perturbations on the exercise values at the exercise times $E_n(x, y, T_n)$, $1 \leq n \leq N-1$, *assuming the volatility is unaffected by the bumps.* We have

$$E_n(x, y, T_n) = 1 - P(x, y, T_n, T_N) - c \sum_{j=n+1}^{N} P(x, y, T_n, T_j) \tau_j,$$

where $P(x, y, t, T)$ is given by (15.1). Let us denote

$$\theta_k(n) = \theta_k(T_n).$$

We have

$$\begin{aligned}
\partial_{\theta_k} P(x, y, T_n, T_j) &= \frac{\theta_k(j)}{P(0, T_n)} \exp\left(-G(t, T)x - \frac{1}{2} G(t, T)^2 y\right) \\
&\quad - \frac{\theta_k(n) P(0, T_j)}{P^2(0, T_n)} \exp\left(-G(t, T)x - \frac{1}{2} G(t, T)^2 y\right) \\
&= \left(\frac{\theta_k(j)}{P(0, T_j)} - \frac{\theta_k(n)}{P(0, T_n)}\right) P(x, y, T_n, T_j).
\end{aligned}$$

thus

$$\begin{aligned}
\partial_{\theta_k} E_n(x, y, T_n) &= -\left(\frac{\theta_k(N)}{P(0, T_N)} - \frac{\theta_k(n)}{P(0, T_n)}\right) P(x, y, T_n, T_N) \\
&\quad - c \sum_{j=n+1}^{N} \left(\frac{\theta_k(j)}{P(0, T_j)} - \frac{\theta_k(n)}{P(0, T_n)}\right) P(x, y, T_n, T_j) \tau_j.
\end{aligned}$$

Similar analysis can be performed for the discount rates

$$F(x, y, T_n, T_{n+1}) = -\frac{1}{\tau_{n+1}} \log P(x, y, T_n, T_{n+1}).$$

In particular,

$$\partial_{\theta_k} F\left(x,y,T_n,T_{n+1}\right) = -\frac{1}{\tau_{n+1} P\left(x,y,T_n,T_{n+1}\right)} \partial_{\theta_k} P\left(x,y,T_n,T_{n+1}\right) = -\frac{1}{\tau_{n+1}} \left(\frac{\theta_k\left(n+1\right)}{P\left(0,T_{n+1}\right)} - \frac{\theta_k\left(n\right)}{P\left(0,T_n\right)} \right). \tag{15.6}$$

Once the derivatives of exercise values and discount rates have been computed, the deltas $\partial_{\theta_k} H_0\left(0\right)$ can be computed via integration as in (10.2) in Theorem 10.1:

$$\partial_\theta H_0\left(T_0\right) = \sum_{n=1}^{N-1} \int dx \int dy \times 1_{\{R_n\}}\left(x,y\right) \times \partial_\theta E_n\left(x,y,T_n\right) \times \psi\left(T_n,x,y\right) - \sum_{n=0}^{N-2} \tau_{n+1} \int dx \int dy \times 1_{\{R_n^c\}}\left(x,y\right) \times \partial_\theta F\left(x,y,T_n,T_{n+1}\right) \times \psi\left(T_n,x,y\right). \tag{15.7}$$

We do not copy the formula here as it can be used directly as written in the theorem, with ν being the Lebesgue measure.

15.4. **Approximate deltas to initial interest rate curve shocks.** Formulas for deltas developed in the previous section require $2\left(N-1\right)$ integrations for each shock θ_k. As explained in Section 10.2, there is a lot of room of reducing the computation effort. Let us try to estimate relative magnitudes of different $\partial_{\theta_k} E_n\left(x,y,T_n\right)$ for a given shock θ_k.

It is clear that for $k<n$,

$$\partial_{\theta_k} E_n\left(x,y,T_n\right) = 0.$$

For $k=n$,

$$\partial_{\theta_k} E_n\left(x,y,T_n\right) = P^{-1}\left(0,T_n\right) P\left(x,y,T_n,T_N\right) + c P^{-1}\left(0,T_n\right) \sum_{j=n+1}^{N} P\left(x,y,T_n,T_j\right) \tau_j. \tag{15.8}$$

For $k=n+1$,

$$\begin{aligned} \partial_{\theta_k} E_n\left(x,y,T_n\right) &= \theta_{n+1}\left(n\right) P^{-1}\left(0,T_n\right) P\left(x,y,T_n,T_N\right) \\ &\quad - c \sum_{j=n+1}^{N} \left(\frac{\theta_{n+1}\left(j\right)}{P\left(0,T_j\right)} - \frac{\theta_{n+1}\left(n\right)}{P\left(0,T_n\right)} \right) P\left(x,y,T_n,T_j\right) \tau_j \\ &= -c P^{-1}\left(0,T_n\right) P\left(x,y,T_n,T_N\right) \\ &\quad - c^2 P^{-1}\left(0,T_n\right) \sum_{j=n+1}^{N} P\left(x,y,T_n,T_j\right) \tau_j \\ &\quad - c \frac{1}{P\left(0,T_{n+1}\right)} P\left(x,y,T_n,T_{n+1}\right) \tau_{n+1}. \end{aligned} \tag{15.9}$$

We see that $\partial_{\theta_{n+1}} E_n\left(x,y,T_n\right)$ is of order c, where as $\partial_{\theta_n} E_n\left(x,y,T_n\right)$ is of order 1. It is not hard to convince oneself that for $k \geq n$, $\partial_{\theta_k} E_n\left(x,y,T_n\right)$ is of order c. The value c is a fixed rate on a swap and is usually of the order 5%. So the shock θ_k, the exercise value $E_k\left(T_k\right)$ is affected the most, $E_n\left(T_n\right)$ are not affected at all for $n>k$, and $E_n\left(T_n\right)$ are affected little for $n < k$. This is in line with the original intention in constructing the shocks expressed in (15.5).

Similar analysis can be performed for the discount rates $F(x, y, T_n, T_{n+1})$.It is clear from formulas (15.6) that for $k = n$ and $k = n + 1$, the expression inside the brackets is of order 1, and for all other k it is of order c.

Our analysis of the magnitude of different $\partial_{\theta_k} E_n(x, y, T_n)$ and $\partial_{\theta_k} F(x, y, T_n, T_{n+1})$ shows that three term dominate for a given shock $\theta_k(\cdot)$, namely $\partial_{\theta_k} E_k(x, y, T_k)$, $\partial_{\theta_k} F(x, y, T_k, T_{k+1})$, $\partial_{\theta_k} F(x, y, T_{k-1}, T_k)$. Therefore, with caution, one can use a simplified formula requiring only "one and a half" integration per delta (we call these three integrations one and a half because each integration is on a reduced domain, either R_n or R_n^c):

$$
\begin{aligned}
\partial_{\theta_k} H_0(T_0) \approx & \int dx \int dy \times 1_{\{R_k\}}(x, y) \times \partial_{\theta_k} E_k(x, y, T_k) \times \psi(T_k, x, y) \\
& -\tau_{k+1} \int dx \times \int dy 1_{\{R_k^c\}}(x) \times \partial_{\theta_k} F(x, T_k, T_{k+1}) \times \psi(x; T_k) \\
& -\tau_k \int dx \int dy \times 1_{\{R_{k-1}^c\}}(x) \times \partial_{\theta_k} F(x, T_{k-1}, T_k) \times \psi(x; T_{k-1}).
\end{aligned}
\tag{15.10}
$$

Another, more accurate approximation for the deltas that has the same computational cost can be obtained as follows. It follows from (15.8) and (15.9) that

$$
\begin{aligned}
\partial_{\theta_{n+1}} E_n(x, y, T_n) &= -c \cdot \partial_{\theta_n} E_n(x, y, T_n) \\
& \quad -c \frac{1}{P(0, T_{n+1})} P(x, y, T_n, T_{n+1}) \tau_{n+1}.
\end{aligned}
$$

Instead of using the approximation

$$
\partial_{\theta_{n+1}} E_n(x, y, T_n) \approx 0,
$$

as in the formula (15.10), a more accurate one can be used

$$
\partial_{\theta_{n+1}} E_n(x, y, T_n) \approx -c \cdot \partial_{\theta_n} E_n(x, y, T_n).
$$

Then the integral of $\partial_{\theta_k} E_n(x, y, T_n)$ can be computed while evaluating $\partial_{\theta_{n+1}} H_0(0)$, and re-used in computing $\partial_{\theta_{n+1}} H_0(0)$. For general k, n, $k < n$, an approximation

$$
\partial_{\theta_k} E_n(x, y, T_n) \approx \theta_k(n) \cdot \partial_{\theta_n} E_n(x, y, T_n)
$$

can be used. Similar formulas can be derived for $\partial_{\theta_k} F(x, T_n, T_{n+1})$. We do not pursue this line further in this paper.

Alternatively, the integrals can be computed one at a time in an adaptive scheme that stops once a contribution of new terms falls below a certain threshold.

We should remember that the formulas (15.7) and (15.10) are computed under the assumption that the volatility process is not affected by the shocks in rates. The volatility in the Cheyette model does however depend on rates. In the next section Cheyette vegas are considered, and in the section after that, volatility contributions to the deltas are considered.

15.5. **Vegas.** The volatility function

$$
\begin{aligned}
\eta(t, x(t), y(t)) &= \lambda(t) \gamma(x(t), y(t)), \\
\lambda(t) &= v_k, \quad t \in (T_k, T_{k+1}],
\end{aligned}
$$

is parametrized by $N-1$ numbers v_k, $0 \leq k \leq N-2$. Model vegas are defined as sensitivities to the shocks of v_k's.

For the purposes of Bermuda swaption valuation, these parameters are usually chosen to match the Black volatilities of core swaptions, i.e. swaptions with expiries T_n and covering a period $[T_n, T_N]$, for $n = 1, \ldots, N$. Their Black volatilities are denoted by u_b^*. There are as many input parameters v_k as there are output parameters u_k^*. Therefore the model satisfies the condition of completeness as defined in Definition 12.1 (at least as long as u_k^* are "reasonable", i.e. do not decline too fast).

It is also easy to see that these parameters satisfy the definition of locality as given in Definition 12.2. A change in v_k for a particular k changes the coefficients of the pricing equation (15.3), (15.4) only on the interval $[T_{k-1}, T_k]$.

A derivative with respect to $v_i.is$ denoted by ∇_i.

As explained in Section 12, the model vega $\nabla_n H_n(x, y, T_n)$ must be computed numerically by rolling the payoff

$$\max\left(H_{n+1}(x, y, T_{n+1}), E_{n+1}(x, y, T_{n+1})\right)$$

in the partial differential equation (15.3) from T_{n+1} to T_n with a shocked volatility $v_n{+}bump$. Once it is computed however, the formula

$$\nabla_n H_0(T_0) = \int dx \int dy \times 1_{\{R_n^c\}}(x, y) \times \nabla_n H_n(x, y, T_n) \times \psi(T_n, x, y) \tag{15.11}$$

can be used.

15.6. **Volatility adjustment for deltas.** A bump to the initial interest rate curve does not only affect H_0 through its effect on E_n, it also changes the volatility in the equation (15.3), which indirectly affects H_0. The first derivative can be computed as a sum of the contributions of these two effects (with the other effect "turned off"), the fact used extensively throughout. The contribution of E_n to the deltas are computed according to the formulas (15.7) or (15.10). In this section the problem of computing the portion of delta that comes from rate dependence of volatility is considered. Recall the volatility specification (15.2):

$$\begin{aligned}
\eta(t, x(t), y(t)) &= \lambda(t)\,\gamma(t, x(t), y(t)), \\
\gamma(t, x(t), y(t)) &= F_{k+1}^{\alpha}(t), \quad t \in (T_k, T_{k+1}], \\
F_k(t) &= \frac{P(t, T_k) - P(t, T_N)}{(T_N - T_k)\,P(t, T_N)}.
\end{aligned}$$

In the notations of Section 13, we set

$$v(t, \cdot) = \gamma(t, x(t), y(t)) = F_{k+1}^{\alpha}(t), \quad t \in (T_k, T_{k+1}].$$

The volatility process is local in v because a shock to each $F_{k+1}(t)$ only affects the volatility on the interval $[T_k, T_{k+1}]$.

Let us fix a shock $\theta(\cdot)$ to the initial interest rate curve (say it is equal to one of the θ_k from Section 15.3). Just like for the vegas, the quantity $\nabla_n^{\theta} H_n(x, y, T_n)$, the impact of θ through volatility on the hold value H_n, must be computed numerically by rolling the payoff

$$\max\left(H_{n+1}(x, y, T_{n+1}), E_{n+1}(x, y, T_{n+1})\right)$$

in the partial differential equation (15.3) from T_{n+1} to T_n with a bumped $F_{n+1}(t)$ appearing in the volatility. Once it is computed, the part of delta attributable to volatility bump on the period $(T_n, T_{n+1}]$ is computed by integration against the survival density

$$\nabla_n^\theta H_0(T_0) = \int dx \int dy \times 1_{\{R_n^c\}}(x,y) \times \nabla_n^\theta H_n(x,y,T_n) \times \psi(T_n,x,y). \tag{15.12}$$

A sum of these terms, one for each n, should be added to $\partial_\theta H_0(T_0)$ as computed in (15.7) or (15.10).

To save on some computations, we can assume that

$$\nabla_n^\theta H_n(x,y,T_n) \approx \alpha_n \times \nabla_n H_n(x,y,T_n)$$

for some coefficient α_n obtained either numerically or heuristically. With this approximation we can then re-use integrals (15.11) computed during vega calculations in (15.12).

15.7. **Volatility calibration adjustment for deltas.** As explained in Section 14, the deltas to interest rate shocks under the assumption that the *market* volatilities are constant is not the same thing as the deltas to interest rate shocks under the assumption that the *model* volatilities are constant. The former can be computed from the latter if we know the derivatives of the market volatilities with respect to the model ones and with respect to the interest rate shocks, see Section 14.

To calibrate the Cheyette model to the market's Black volatilities of European swaptions, we need to express these volatilities as functions of model volatilities $\lambda(\cdot)$. These functions must be fast to compute. The paper [And01b] suggests the following approximation. Let $S(x,y,t)$ be a swap rate, observed at time t and expressed as a function of state variables x and y, of a particular European swaption with maturity T whose Black's volatility needs to be expressed as a function of model volatilities $\lambda(\cdot)$. In the swap measure associated with this swaption, $S(x(t), y(t), t)$ is a martingale. Therefore, it satisfies the SDE,

$$dS(x(t), y(t), t) = \left[\frac{\partial}{\partial x} S(x(t), y(t), t)\right] \lambda(t)\gamma(t, x(t), y(t))\, d\tilde{W}(t),$$

(here $\tilde{W}$ is a Brownian motion in the swap measure). Therefore,

$$\mathrm{Var} S(x(T), y(T), T) = \mathbf{E} \int_0^T \left[\frac{\partial}{\partial x} S(x(t), y(t), t)\right]^2 \lambda^2(t)\gamma^2(x(t), y(t))\, dt.$$

Evaluating the right-hand side at $x(t) = x(0) = 0$, $y(t) = y(0) = 0$, we obtain an approximate expression

$$\mathrm{Var} S(x(T), y(T), T) \approx \int_0^T \left[\frac{\partial}{\partial x} S(x,y,t)\right]^2 \Bigg|_{x=y=0} \lambda^2(t)\gamma^2(0,0)\, dt,$$

for the variance of the swap rate. The Black's term volatility for this swaption can then be approximated by

$$\sigma_S = S^{-1}(0,0,0) \sqrt{\frac{1}{T}\int_0^T \left[\frac{\partial}{\partial x} S(x,y,t)\right]^2 \Bigg|_{x=y=0} \lambda^2(t)\gamma^2(0,0)\, dt}.$$

This expression can be differentiated with respect to the model's volatilities $\lambda^2(t)$ and interest rate shocks θ (they will affect $S(0,0,0)$ and $\gamma^2(t,0,0)$). The derivatives can then be used in the formula (14.3) to obtain the deltas to interest rate curve shocks while keeping the *market* volatilities constant.

16. Computing deltas of Bermuda swaptions in Monte-Carlo simulation

Methods for valuing Bermuda swaptions in Monte-Carlo simulation have been developed relatively recently, see [And99] and [LS98]. These advances allow one to price Bermuda swaptions in models with much richer volatility structure than was previously available with low-dimensional Markovian models. As such, the models can be calibrated to a much broader set of market instruments (in the context of Bermuda swaption pricing, these usually include European swaptions and possibly interest rate caps).

The representations of deltas developed in previous sections allow us to speed up their calculation not only in a PDE rollback, but also in Monte-Carlo simulation as well. This section explores this direction.

A similar question, for general interest rate instruments, is addressed in [GZ99]. Our focus on Bermuda swaptions allows us to derive, we believe, more useful formulas for this particular class of interest rate derivatives.

When talking about pricing Bermuda swaptions in Monte-Carlo simulation, a model of choice is usually a BGM, or Libor Market, model (see [MR97, Chapter 14]). This model can be calibrated to the whole European swaption grid (and all interest rate caps if required), thus incorporating essentially all market information relevant to pricing Bermuda swaptions.

16.1. **The BGM model.** There are many flavors of BGM models, we concentrate on a particular kind. The results we derive are not specific to this choice.

Libor rates are the subject of modeling in BGM models, and the dynamics of the interest rate curve is expressed in their terms. We use Libor rates that align with the tenor structure $\{T_n\}_{n=0}^{N}$ that we have been working with throughout the paper. This is purely for notational convenience, as the driving Libor rates can be based on any tenor structure.

We define the n-th forward Libor rate $L_n(t)$ by the usual formula

$$L_n(t) = \frac{P(t, T_n) - P(t, T_{n+1})}{\tau_n P(t, T_{n+1}).}, \quad 0 \leq n \leq N.$$

We impose the following dynamics on each of the forward Libor rates,

$$\frac{dL_n(t)}{L_n(t)} = g_n(t)\; dW^{T_{n+1}}(t), \quad n = 1, \ldots, N, \quad t \in [0, T_n].$$

Here $g_n(\cdot)$ is a deterministic function $\mathbb{R}_+ \to \mathbb{R}^n$, and $W^{T_{n+1}}(\cdot)$ is an K-dimensional Brownian motion under the T_{n+1}-forward measure. It is known (see [MR97]) that this specification leads to a valid HJM model as defined in (3.1).

A numeraire different from the money-market account is usually chosen. A discrete money-market numeraire N_t is defined by

$$\begin{aligned} N_{T_0} &= 1, \\ N_{T_{n+1}} &= N_{T_n} \times (1 + \tau_n L_n(T_n)), \quad 1 \leq n \leq N, \\ N_t &= P(t, T_{n+1})\, N_{T_{n+1}}, \quad t \in [T_n, T_{n+1}]. \end{aligned}$$

The dynamics of all forward Libor rates under the same measure, the measure associated with N_t, are given by

(16.1)
$$\frac{dL_n(t)}{L_n(t)} = \left\langle g_n(t), \sum_{j=1}^{n} 1_{\{t<T_j\}} \frac{\tau_j L_j(t)}{1+\tau_j L_j(t)} g_k(t) \right\rangle dt + g_n(t)\, dW(t), \quad n = 1, \ldots, N,$$

where $\langle \cdot, \cdot \rangle$ denotes scalar product, and $W(\cdot)$ is a Brownian motion under this measure.

16.2. **Valuing Bermuda swaptions in the BGM model.** Monte-Carlo valuation of a Bermuda swaption produces a lower bound for the price (recently, methods for computing an upped bound became available, see [AB01]). Let us review the basics of Monte-Carlo valuation of Bermuda swaptions. A typical lower bound for a Bermuda swaption price is computed according to the following procedure.

1. A number of Monte-Carlo paths (pre-simulation) is generated;
2. Based on these paths, an exercise strategy for Bermuda swaption is estimated;
3. More Monte-Carlo paths are simulated, and for each path, the Bermuda swaption is exercised according to the exercise strategy from the previous step.

The exercise strategy in Step 2 is usually chosen to be optimal (given the pre-simulated paths) in the pre-specified subclass of strategies. This procedure returns a lower bound because the exercise strategy used, while optimal it the subclass, is necessarily sub-optimal in the class of **all** exercise strategies.

There are two general approaches to choosing the exercise strategy – Andersen's (see [And99]) and Longstaff-Schwartz's (see [LS98]). Andersen's idea involves expressing the exercise boundary in terms of the underlying instruments, swaps and swaptions. Then the free parameters of the exercise boundary are optimized numerically, one exercise at a time, in a backward induction. Longstaff and Schwartz propose defining some regression variables for each exercise time, and regressing the hold value and the exercise value on those variables in a backward induction.

16.3. **Deltas as knock-out contingent claims.** The details of the methods described in the previous section are unimportant to us. What is important, however, is that both methods produce an estimate of the exercise region R_n (and, correspondingly, the hold region R_n^c) for each n, $1 \le n \le N-1$. Each R_n is expressed in terms of the variables available in forward Monte-Carlo simulation at time T_n.

For example, a very simple exercise strategy may say that we should exercise on the first n when $E_n(T_n, \omega)$ is greater than a pre-determined bound b_n (here we regard ω as a simulated path of a Monte-Carlo). In this case, R_n is defined as

$$R_n = 1_{\{E_n(T_n,\omega)>b_n\}}.$$

Recall Corollary 8.2 that expresses deltas as values of certain knock-out contingent claims. The knockout condition is expressed in terms of R_n. After choosing the exercise strategy in Step 2 in the algorithm described in the previous section, estimates of R_n for each n are available to us. Therefore, the knockout condition can always be evaluated in a forward Monte-Carlo simulation. Hence, the deltas can be evaluated in the same simulation as Monte-Carlo – if we know the payoff of said knock-out contingent claims.

The same idea can be expressed in terms of Corollary 8.1. The estimate of the optimal exercise time index η from the corollary is available during the Step 3 of the valuation algorithm. Therefore, for each simulated path ω, $\eta(\omega)$ can be evaluated. The underlying deltas can then be estimated by (ω_i, $i = 1, \ldots, I$ are Monte-Carlo paths in Step 3)

$$\Delta_n^u h_0(A) \approx \frac{1}{I} \sum_{i=1}^{I} \left(N_{\eta(\omega_i)} \times 1_{\{\eta(\omega_i)=n\}} \times d_n^u(\omega_i) \right), \quad 1 \leq n \leq N-1,$$

with a similar expression holding for the discount deltas.

The only unknown in this formula is $d_n^u(\omega_i)$, which is the shock to the $E_n(T_n)$ induced by some shock to the initial interest rate curve.

16.4. **Initial interest rate curve shocks and their effect on exercise values.** Recall our approach to computing market deltas. A shock θ to the initial interest rate curve is chosen. The effect of this shock on the exercise values $E_n(T_n)$ and discount rates $F(T_n, T_{n+1})$ at exercise times is computed. Then the resulting shocks to $E_n(T_n)$ and $F(T_n, T_{n+1})$ are integrated against the survival density.

The integration step can be performed in a Monte-Carlo simulation, as explained above. The remaining piece of the puzzle, how to propagate initial interest rate curve shocks to future exercise values (and discount rates) is discussed in this section.

We follow closely follow the ideas of [GZ99].

The BGM model is expressed in terms of the Libor rates. It is natural to think of shocks in terms of Libor rates as well.

Let θ_k be a shock that affects the k-th Libor rate (at time 0), $L_k(0)$, only. We denote the derivative with resect to that shock by $\partial_k \triangleq \partial_{\theta_k}$.

The equation for L_n is

$$dL_n(t) = L_n(t) \left\langle g_n(t), \sum_{j=1}^{n} 1_{\{t<T_j\}} \frac{\tau_j L_j(t)}{1+\tau_j L_j(t)} g_k(t) \right\rangle dt + L_n(t) g_n(t)\, dW(t).$$

Differentiating this equation through,

$$\begin{aligned} d(\partial_k L_n)(t) &= (\partial_k L_n)(t) \left\langle g_n(t), \sum_{j=1}^{n} 1_{\{t<T_j\}} \frac{\tau_j L_j(t)}{1+\tau_j L_j(t)} g_k(t) \right\rangle dt \\ &\quad + L_n(t) \left\langle g_n(t), \sum_{j=1}^{n} 1_{\{t<T_j\}} \frac{\tau_j (\partial_k L_j)(t)}{(1+\tau_j L_j(t))^2} g_k(t) \right\rangle dt \\ &\quad + (\partial_k L_n)(t) g_n(t)\, dW(t), \end{aligned} \tag{16.2}$$

a stochastic differential equation for $\partial_k L_n$ is obtained, The initial condition for the SDE is given by

$$(\partial_k L_n)(0) = 1_{\{k=n\}}.$$

The equations (16.2) for $\partial_k L_n$ for all k, n, can be simulated in the same Monte-Carlo run in which the equations (16.1) for L_n are simulated. Conceptually it is no more difficult, and technically it is not much slower that simulating the L_n's only, because all the quantities involved in (16.2) are also needed in (16.1).

A simplified version of these equations can be derived by ignoring the drifts for the dynamics of the deltas. The following SDE then holds (approximately) for $k = n$,

$$\begin{aligned} (\partial_n L_n)(t) &= 1 + \int (\partial_n L_n)(t)\, g_n(t)\, dW(t), \\ (\partial_n L_n)(t) &= \frac{L_n(t)}{L_n(0)}, \end{aligned}$$

and for $k \neq n$,

$$\begin{aligned} (\partial_k L_n)(t) &= \int (\partial_k L_n)(t)\, g_n(t)\, dW(t), \\ (\partial_k L_n)(t) &\equiv 0. \end{aligned}$$

Finally we observe that the values $E_n(T_n)$ can be expressed as functions of $L_k(T_n)$ for $k \geq n$. The same is true for the discount rates $F(T_n, T_{n+1})$. Note that for each n, $F(T_n, T_{n+1})$ is a continuous-compounding analog of a simple-compounding rate $L_n(T_n)$. Using $\partial_k F(T_n, T_{n+1}) = (\partial_k L_n)(T_n)$ is likely to be accurate enough.

16.5. **Computing deltas to shocks of the initial interest rate curve.** The algorithm for computing all market deltas $\partial_k H_0(T_0)$ in the same simulation as the value $H_0(T_0)$ works as follows. First, a number of pre-simulation paths is run and exercise regions for each T_n are estimated, just like in the normal Monte-Carlo valuation of a Bermuda swaption. Then, in the post-simulation, for each path ω,

1. All $(\partial_k L_n)(t, \omega)$ are computed alongside with all $L_n(t, \omega)$ using equations (16.2);
2. The estimate of optimal exercise time index $\eta(\omega)$ is set to be the first n for which the path ω hits the exercise region R_n;
3. The values $\partial_k E_{\eta(\omega)}(T_n, \omega)$ and $\partial_k F(T_n, T_{n+1})$ for each k are computed by expressing them in terms of $\partial_k L_{\eta(\omega)}(t, \omega)$.

According to Theorem 10.1, the estimate of each market delta $\partial_k H_0(T_0)$, $1 \leq k \leq N-1$, can be computed as

$$\begin{aligned} \partial_k H_0(T_0) \approx\ & \frac{1}{I}\sum_{i=1}^{I} \left(N_{\eta(\omega_i)} \times \partial_k E_{\eta(\omega)}(T_n, \omega)\right) \\ & - \frac{1}{I}\sum_{i=1}^{I} \sum_{n=1}^{\eta(\omega_i)-1} \left(N_n \times \tau_{n+1} \times \partial_k F(T_n, T_{n+1})\right). \end{aligned} \tag{16.3}$$

16.6. **Vegas.** Computing vegas (volatility sensitivities) of Bermuda swaptions in Monte-Carlo simulation is much more complicated. We recall from Section 12 that the quantities $\nabla_n H_n(T_n)$ should be computed by rolling back the appropriate payoff from T_{n+1} to T_n. Performing rollback is rather tricky in a Monte-Carlo simulation. It is not completely impossible however. The Longstaff-Schwartz framework is, in fact, based on computing

$$H_n(T_n) = \pi_{T_n} \max\left(H_{n+1}(T_{n+1}), E_{n+1}(T_{n+1})\right)$$

in a rollback procedure by regressing $H_n(T_n)$ in the pre-simulation step on a set of "explanatory" variables. The same "regression rollback" can be used in post-simulation as well, thus giving us a way to compute $\nabla_n H_n(T_n)$ in simulation. Full details of this procedure are beyond the scope of this paper. Once $\nabla_n H_n(T_n)$ are computed, the vegas $\nabla_n H_0(T_0)$ can computed in a way similar to deltas, by averaging out $\nabla_n H_n(T_n)$ over those paths ω for which $\eta(\omega) = n$.

We note that the volatility process for the BGM model satisfies our definition of locality. Typically $g_n(t)$ are chosen to be piecewise constant on time intervals $[T_k, T_{k+1})$. So the parameter v_k (notations from Section 12) is equal to $\{g_n(T_k), 1 \le n \le N\}$. Recall that each $g_n(T_k)$ is multi-dimensional as well.

Mapping of model vegas to market vegas can be done with the aid of the approximation formulas for Black volatility of European swaptions (see e.g. [AA98]).

16.7. **Volatility adjustment for deltas.** No volatility adjustment for deltas is needed. The deltas computed in Section 16.5 already include volatility effects. Even though the volatility process for the model does depend on interest rates, this dependence was taken into account in the formula (16.3). This is so because the volatilities $g_n(t) F_n(t)$ were also shocked in the equation (16.2), as clear from the last term.

References

[AA98] Leif B.G. Andersen and Jesper Andreasen. Volatility skews and extensions of the Libor Market Model. Working paper, 1998.

[AA00] Leif B.G. Andersen and Jesper Andreasen. Factor dependence of Bermudan swaption prices: Fact or fiction? Working Paper, 2000.

[AB01] Leif B.G. Andersen and Mark Broadie. A primal-dual simulation algorithm for pricing of multi-dimensional options. Working Paper, 2001.

[ABR01] Leif B.G. Andersen and Rupert Brotherton-Ratcliffe. Extended libor market models with stochastic volatility. Working paper, 2001.

[And96] Jesper Andreasen. *Essays on Contingent Pricing*. PhD thesis, University of Aarhus, 1996.

[And99] Leif B.G. Andersen. A simple approach to the pricing of Bermudan swaptions in the multi-factor Libor Market Model. Gen Re working paper, 1999.

[And01a] Jesper Andreasen. The pricing of Bermuda swaptions. Risk conference, Paris, 2001.

[And01b] Jesper Andreasen. Turbo-charging the Cheyette model. Bank of America working paper, 2001.

[Ber99] H.-P. Bermin. A general approach to hedging options: Applications to barrier and partial barrier options. Working paper, Lund University, Lund, Sweden, 1999.

[BL94] P.P. Boyle and S.H. Lau. Bumping up against the barrier with the binomial method. *Journal of Derivatives*, 1:6–14, 1994.

[Car01] Peter Carr. Deriving derivatives of derivative securities. *Journal of Computational Finance*, 2000/2001.

[Che91] O. Cheyette. Markov representation of the Heath-Jarrow-Morton model. Working paper, BARRA, 1991.

[DJ97] M.A.H. Dempster and J.P.Hutton. Fast numerical valuation of American, exotic and complex options. Working paper, 1997.

[Dod02] Stephen Dodds. Pricing and hedging Bermudan swaptions with the BGM model. Barclays working paper, 2002.

[FLLT99] Eric Fournie, Jean-Michel Lebuchoux, Pierre-Louis Lions, and Nizar Touzi. Application of Malliavin calculus to Monte Carlo methods in finance. *Finance and Stochastics*, 3:391–412, 1999.

[GZ99] Paul Glasserman and Xiaoling Zhao. Fast greeks in forward Libor models. Working paper, 1999.

[Jam] Leland Jameson. On the wavelet optimized finite difference method. Working paper, NASA Langley research center.

[JK98] Phil J.Hunt and J.E. Kennedy. Implied interest rate pricing models. *Finance and Stochastics*, 2(3), 1998.

[LS98] Francis Longstaff and Eduardo Schwartz. Valuing american options by simulation: A simple least-squares approach. Working paper, The Anderson School, UCLA, 1998.

[LSSC99] Francis Longstaff, Eduardo Schwartz, and E. Santa-Clara. Throwing away a billion dollars: The cost of suboptimal exercise in the swaptions market. Working paper, The Anderson School, UCLA, 1999.

[MM94] K. Morton and D. Mayers. *Numerical Solution of Partial Differential Equations.* Cambridge University Press, 1994.

[MR97] Marek Musiela and Marek Rutkowski. *Martingale Methods in Financial Modeling.* Springer, 1997.

[RE98] L.C.G Rogers and E.J.Stapleton. Fast accurate binomial pricing. *Finance and Stochastics*, 2(1), 1998.

[RR] Mark Rubinstein and Eric Reiner. Breaking down the barriers. *Risk*, 4(8).

[RS97] G. O. Roberts and C. F. Shortland. Pricing barrier options with time-dependent coefficients. *Mathematical Finance*, 7(1):83–93, January 1997.

Appendix A. Proofs

A.1. **Proof of Lemma 7.1**[1]**.** We have for all $\omega \in \Omega$ that

$$\begin{aligned}&\max(X+\varepsilon D, Y) - \max(X,Y) - \varepsilon 1_{\{X\geq Y\}} D \\ &\qquad = 1_{\{Y-\varepsilon D\leq X<Y\}} \times (X - Y + \varepsilon D) + 1_{\{Y-\varepsilon D > X\geq Y\}} \times (Y - X - \varepsilon D).\end{aligned}$$

Since

$$\begin{aligned} 0 &\leq 1_{\{Y-\varepsilon D\leq X<Y\}} \times (X-Y+\varepsilon D) \leq \varepsilon \times 1_{\{Y-\varepsilon D\leq X<Y\}} \times D, \\ 0 &\leq 1_{\{Y-\varepsilon D> X\geq Y\}} \times (Y-X-\varepsilon D) \leq -\varepsilon \times 1_{\{Y-\varepsilon D> X\geq Y\}} \times D, \end{aligned}$$

we get,

$$\begin{aligned}0 \leq \frac{1}{\varepsilon}&\left[\max(X+\varepsilon D, Y) - \max(X,Y) - \varepsilon 1_{\{X>Y\}} D\right] \\ &\qquad\qquad \leq \left(1_{\{Y-\varepsilon D\leq X<Y\}} - 1_{\{Y-\varepsilon D> X\geq Y\}}\right) \times D.\end{aligned}$$

Clearly

$$\begin{aligned} 1_{\{Y-\varepsilon D\leq X<Y\}} &\to 1_{\{X=Y\}} \times 1_{\{D>0\}} \quad \text{as} \quad \varepsilon \to 0, \\ 1_{\{Y-\varepsilon D> X\geq Y\}} &\to 1_{\{X=Y\}} \times 1_{\{D<0\}} \quad \text{as} \quad \varepsilon \to 0, \end{aligned} \tag{A.1}$$

almost surely. Also, since

$$\begin{aligned} \left|1_{\{Y-\varepsilon D\leq X<Y\}} \times D\right| &\leq |D|, \\ \left|1_{\{Y-\varepsilon D> X\geq Y\}} \times D\right| &\leq |D|, \end{aligned}$$

and

$$\mathbf{E}|D| < \infty,$$

by the dominated convergence theorem and from (A.1),

$$\begin{aligned} \lim_{\varepsilon\to 0} \mathbf{E}_t\left(1_{\{Y-\varepsilon D\leq X<Y\}} D\right) &= \mathbf{E}_t\left(1_{\{X=Y\}} \times 1_{\{D>0\}}\right), \\ \lim_{\varepsilon\to 0} \mathbf{E}_t\left(1_{\{Y-\varepsilon D> X\geq Y\}} D\right) &= \mathbf{E}_t\left(1_{\{X=Y\}} \times 1_{\{D<0\}}\right). \end{aligned}$$

Since $P(X=Y)=0$, the right-hand sides of the equations above are zero. This, together with (**??**), implies that

$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \mathbf{E}_t\left[\max(X+\varepsilon D, Y) - \max(X,Y) - \varepsilon 1_{\{X>Y\}} D\right] = 0.$$

The proof of the lemma is complete.

A.2. **Proof of Theorem 7.2.** Recall the main recursion for Bermuda swaption values (6.6),

$$h_k(a) = \exp(-\tau_{k+1} f_k)\, \mathbf{E}_{T_k}^{T_{k+1}} \max\{h_{k+1}(a), e_{k+1}\}, \quad k = N-2, \dots, 0.$$

Assume $k > n-1$. Then the shock to the n-th underlying e_n does not affect $h_k(a)$ at all. This is so because $h_k(a)$ depends only on e_i with $i \geq k+1$, as clear from the formula above.

[1]Leif Andersen provided us with a simplified proof of this lemma which we gratefully acknowledge.

Assume $k = n - 1$. Then the shock to the n-th underlying $e_n \equiv e_{k+1}$ obviously affects h_k but does not affect h_i for $i \geq k + 1$. Therefore, we can apply Lemma 7.1 with $X = e_{k+1}$, $Y = h_{k+1}(a)$, $D = d^u_{k+1}$ to obtain

$$\Delta^u_n h_k(a) = \exp(-\tau_{k+1} f_k)\, \mathbf{E}^{T_{k+1}}_{T_k} 1_{\{e_{k+1}(a) \geq h_{k+1}(a)\}} d^u_{k+1},$$

as stated in the theorem.

Finally, let us consider a case $k < n - 1$. Then the shock to the n-th underlying e_n does not affect e_{k+1} (shocks to different e's are independent by definition) but it does affect h_{k+1}. Hence we can apply Lemma 7.1 with $X = h_{k+1}(a)$, $Y = e_{k+1}$, $D = \Delta^u_n h_{k+1}(a)$ to obtain

$$\Delta^u_n h_k(a) = \exp(-\tau_{k+1} f_k)\, \mathbf{E}^{T_{k+1}}_{T_k} 1_{\{e_{k+1}(a) \leq h_{k+1}(a)\}} \Delta^u_n h_{k+1}(a).$$

The theorem is proved.

A.3. **Proof of Theorem 7.3.** The proof is analogous to the proof of Theorem 7.2.

For $k > n$, $h_k(a)$ is unaffected by a shock to the n-th discount rate f_n at all, thus

$$\Delta^d_n h_k(a) = 0.$$

For $k = n$, the expression

$$\mathbf{E}^{T_{k+1}}_{T_k} \max\{h_{k+1}(a), e_{k+1}\}$$

is unaffected by a shock to the n-th discount rate f_n at all, thus

$$\begin{aligned} \Delta^d_k h_k(a) &= \left[\Delta^d_k \exp(-\tau_{k+1} f_k)\right] \times \mathbf{E}^{T_{k+1}}_{T_k} 1_{\{e_{k+1}(a) \leq h_{k+1}(a)\}} \Delta^u_n h_{k+1}(a) \\ &= -\tau_{k+1} d^d_k \times \mathbf{E}^{T_{k+1}}_{T_k} 1_{\{e_{k+1}(a) \leq h_{k+1}(a)\}} \Delta^u_n h_{k+1}(a). \end{aligned}$$

For $k < n$, we use Lemma 7.1 with $X = h_{k+1}(a)$, $Y = e_k$, $D = \Delta^d_n h_{k+1}(a)$ to obtain

$$\Delta^d_n h_k(a) = \exp(-\tau_{k+1} f_k)\, \mathbf{E}^{T_{k+1}}_{T_k} 1_{\{e_{k+1}(a) \leq h_{k+1}(a)\}} \Delta^d_n h_{k+1}(a).$$

The theorem is proved.

A.4. **Proof of Theorem 9.2.** Assume

$$T_n < t \leq T_{n+1}.$$

Then

$$\begin{aligned} \Psi(Y; t) &= \pi_0\left(1_{\{\eta > n\}} \times 1_{\{x(t) \in Y\}}\right) \\ &= \mathbf{E}_0\left(e^{-\int_0^t r(u)\,du} \times 1_{\{\eta > n\}} \times 1_{\{x(t) \in Y\}}\right) \\ &= \mathbf{E}_0\left(\mathbf{E}\left(e^{-\int_0^t r(u)\,du} \times 1_{\{\eta > n\}} \times 1_{\{x(t) \in Y\}} \middle| x(T_n)\right)\right) \\ &= \mathbf{E}_0\left(e^{-\int_0^{T_n} r(u)\,du} \times 1_{\{\eta > n\}} \times \mathbf{E}\left(e^{-\int_{T_n}^t r(u)\,du} \times 1_{\{x(t) \in Y\}} \middle| x(T_n)\right)\right) \\ &= \mathbf{E}_0\left(e^{-\int_0^{T_n} r(u)\,du} \times 1_{\{\eta > n\}} \times \Phi_{T_n, x(T_n)}(Y; t)\right). \end{aligned}$$

In the line next to last we used the fact that η is a stopping time and this the event $\{\eta > n\}$ belongs to the sigma-algebra $\mathcal{H}_{T_n}$.

From this formula we obtain

$$\psi(y; t) = \mathbf{E}_0\left(e^{-\int_0^{T_n} r(u)\,du} \times 1_{\{\eta > n\}} \times \phi_{T_n, x(T_n)}(y; t)\right).$$

Differentiating this equality with respect to t, exchanging the order of differentiation and taking the expectation, applying (9.1) and exchanging the order of the linear operator $\Lambda^* - r(t)$ and the expectation operator, we obtain that the same equation as (9.1) holds for $\psi(y;t)$,

$$\frac{\partial}{\partial t}\psi(y;t) = (\Lambda^*\psi)(y;t) - r(t)\psi(y;t)$$

for

$$t \in (T_n, T_{n+1}].$$

This equation governs the dynamics of the density $\psi(y;t)$ for t up until the time T_{n+1}. What happens to $\psi(y;t)$ when t "crosses over" T_{n+1}? Let

$$t = T_{n+1} + \varepsilon,$$

where $\varepsilon > 0$. Then

$$\begin{aligned}\Psi(Y; T_{n+1}+\varepsilon) &= \pi_0\left(1_{\{\eta>n+1\}} \times 1_{\{x(T_{n+1}+\varepsilon)\in Y\}}\right) \\ &= \pi_0\left(1_{\{\eta>n\}} \times 1_{\left\{x(T_{n+1})\in R^c_{n+1}\right\}} \times 1_{\{x(T_{n+1}+\varepsilon)\in Y\}}\right).\end{aligned}$$

By the continuity of the trajectories of x we have for closed sets Y,

$$\lim_{\varepsilon\downarrow 0} 1_{\{x(T_{n+1}+\varepsilon)\in Y\}} = 1_{\{x(T_{n+1})\in Y\}}.$$

By the dominated convergence theorem,

$$\lim_{\varepsilon\downarrow 0}\Psi(Y; T_{n+1}+\varepsilon) = \pi_0\left(1_{\{\eta>n\}} \times 1_{\left\{x(T_{n+1})\in R^c_{n+1}\right\}} \times 1_{\{x(T_{n+1})\in Y\}}\right).$$

By definition, the right-hand side is equal to

$$\Psi\left(Y \cap R^c_{n+1}; T_{n+1}\right).$$

This can be expressed in terms of the survival density as

$$\phi(y; T_{n+1}+0) = \phi(y; T_{n+1}) \times 1_{\left\{y\in R^c_{n+1}\right\}}.$$

This completes the proof of the theorem.

Appendix B. Recursive formulas for Gammas

In this section we derive recursive formulas for bucketed gammas by differentiating recursive formulas for bucketed deltas. Both deltas and gammas are understood as derivatives of $h_n(a)$ in directions d^u_n, $d^d_n \in \mathcal{A}$, $n = 1, \dots, N-1$.

We differentiate recursive formulas for deltas from Theorems 7.2, 7.3 to obtain recursive formulas for gammas. Without loss of generality we assume

$$m \leq n.$$

We use the following analog to Lemma 7.1. Under the appropriate technical conditions,

$$\lim_{\varepsilon\to 0}\varepsilon^{-1}\mathbf{E}_t\left(1_{\{X+\varepsilon D>Y\}} - 1_{\{X>Y\}}\right) = \mathbf{E}_t\left(\delta_{\{X=Y\}}D\right).$$

For "underlying-underlying" bucketed gammas $\Delta^u_m\Delta^u_n$ we have

- For $k > m-1$,

$$\Delta^u_m\Delta^u_n h_k(A) = 0;$$

- For $k = m-1$,

$$\begin{aligned}\Delta_m^u \Delta_n^u h_k(A) &= 1_{\{m=n\}} \times \pi_{T_k}\left(\delta_{\{H_m(T_m)=E_m(T_m)\}} \times \left(d_m^u\right)^2\right) \\ &\quad -1_{\{m\neq n\}} \times \pi_{T_k}\left(\delta_{\{H_m(T_m)=E_m(T_m)\}} \times d_m^u \times \Delta_n^u h_{k+1}(A)\right)\end{aligned}$$

- For $k < m-1$,

$$\begin{aligned}\Delta_m^u \Delta_n^u h_k(A) &= \pi_{T_k}\left(\delta_{\{H_{k+1}(T_{k+1})=E_{k+1}(T_{k+1})\}} \times \Delta_m^u h_{k+1}(A) \times \Delta_n^u h_{k+1}(A)\right) \\ &\quad +\pi_{T_k}\left(1_{\{H_{k+1}(T_{k+1})>E_{k+1}(T_{k+1})\}} \times \Delta_m^u \Delta_n^u h_{k+1}(A)\right).\end{aligned}$$

For "underlying-discount" bucketed gammas $\Delta_m^u \Delta_n^d$ we have

- For $k > m-1$,

$$\Delta_m^u \Delta_n^d h_k(A) = 0;$$

- For $k = m-1$,

$$\Delta_m^u \Delta_n^d h_k(A) = -\pi_{T_k}\left(\delta_{\{H_m(T_m)=E_m(T_m)\}} \times d_m^u \times \Delta_n^d h_{k+1}(A)\right);$$

- For $k < m-1$,

$$\begin{aligned}\Delta_m^u \Delta_n^d h_k(A) = \pi_{T_k}&\left(\delta_{\{H_{k+1}(T_{k+1})=E_{k+1}(T_{k+1})\}} \times \Delta_m^u h_{k+1}(A) \times \Delta_n^d h_{k+1}(A)\right) \\ &+\pi_{T_k}\left(1_{\{H_{k+1}(T_{k+1})>E_{k+1}(T_{k+1})\}} \times \Delta_m^u \Delta_n^d h_{k+1}(A)\right).\end{aligned}$$

- For "discount-underlying" bucketed gammas $\Delta_m^d \Delta_n^u$ we have
- For $k > n-1$ or $k > m$,

$$\Delta_m^d \Delta_n^u h_k(A) = 0;$$

- For $k = m$,

$$\Delta_m^d \Delta_n^u h_k(A) = -\tau_{m+1} \times d_m^d \times \Delta_n^u h_k(A);$$

- For $k = n-1$, $k \neq m$,

$$\Delta_m^d \Delta_n^u h_k(A) = 1_{\{m=n\}} \times \tau_{n+1} \times \pi_{T_k}\left(\delta_{\{H_n(T_n)=E_n(T_n)\}} \times d_n^u \times d_n^d \times H_n(T_n)\right);$$

- For $k < m$,

$$\begin{aligned}\Delta_m^d \Delta_n^u h_k(A) &= \pi_{T_k}\left(\delta_{\{H_{k+1}(T_{k+1})=E_{k+1}(T_{k+1})\}} \times \Delta_m^d h_{k+1}(A) \times \Delta_n^u h_k(A)\right) \\ &\quad +\pi_{T_k}\left(1_{\{H_{k+1}(T_{k+1})>E_{k+1}(T_{k+1})\}} \times \Delta_m^d \Delta_n^u h_{k+1}(A)\right).\end{aligned}$$

For "discount-discount" bucketed gammas $\Delta_m^d \Delta_n^d$ we have

- For $k > m$,

$$\Delta_m^d \Delta_n^d h_k(A) = 0;$$

- For $k = m$,

$$\Delta_m^d \Delta_n^d h_k(A) = -\tau_{m+1} \times d_m^d \times \Delta_n^d H_k;$$

- For $k < m$,

$$\begin{aligned}\Delta_m^d \Delta_n^d h_k(A) &= \pi_{T_k}\left(\delta_{\{H_{k+1}(T_{k+1})=E_{k+1}(T_{k+1})\}} \times \Delta_m^d h_{k+1}(A) \times \Delta_n^d h_{k+1}(A)\right) \\ &\quad +\pi_{T_k}\left(1_{\{H_{k+1}(T_{k+1})>E_{k+1}(T_{k+1})\}} \times \Delta_m^d \Delta_n^d h_{k+1}(A)\right).\end{aligned}$$

These recursive formulas can be stringed together to get expressions for various gammas. For example, for the "underlying-underlying" bucketed gammas $\Delta_m^u \Delta_n^u$ (recall that $m \leq n$),

$$\begin{aligned}
\Delta_m^u \Delta_n^u h_0\left(A\right) &= \sum_{k=1}^{m-1} \pi_0 \left(\left(\prod_{i=1}^{k-1} 1_{\{H_i(T_i) > E_i(T_i)\}} \right) \delta_{\{H_k(T_k) = E_k(T_k)\}} \Delta_m^u H_k\left(T_k\right) \Delta_n^u H_k\left(T_k\right) \right) \\
&\quad + \pi_0 \left(\left(\prod_{i=1}^{m-1} 1_{\{H_i(T_i) > E_i(T_i)\}} \right) \delta_{\{H_m(T_m) = E_m(T_m)\}} d_m^u \left(1_{\{m=n\}} d_m^u - 1_{\{m \neq n\}} \Delta_n^u H_m\left(T_m\right) \right) \right).
\end{aligned}$$

Similar formulas can be derived for other gammas.

233 South Wacker Drive, Suite 2800, Chicago, IL 60606
E-mail address: `vladimir.piterbarg@bankofamerica.edu`