THE UNIVERSITY OF
WARWICK

AUTHOR: **Duy Pham** DEGREE: **Ph.D.**

TITLE: **Markov-functional and Stochastic Volatility modelling**

DATE OF DEPOSIT: .....

I agree that this thesis shall be available in accordance with the regulations governing the University of Warwick theses.

I agree that the summary of this thesis may be submitted for publication.

I **agree** that the thesis may be photocopied (single copies for study purposes only).

Theses with no restriction on photocopying will also be made available to the British Library for microfilming. The British Library may supply copies to individuals or libraries. subject to a statement from them that the copy is supplied for non-publishing purposes. All copies supplied by the British Library will carry the following statement:

> "Attention is drawn to the fact that the copyright of this thesis rests with its author. This copy of the thesis has been supplied on the condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the author's written consent."

AUTHOR'S SIGNATURE: .....

USER'S DECLARATION

1. I undertake not to quote or make use of any information from this thesis without making acknowledgement to the author.
2. I further undertake to allow no-one else to use this thesis while it is in my care.

DATE SIGNATURE ADDRESS

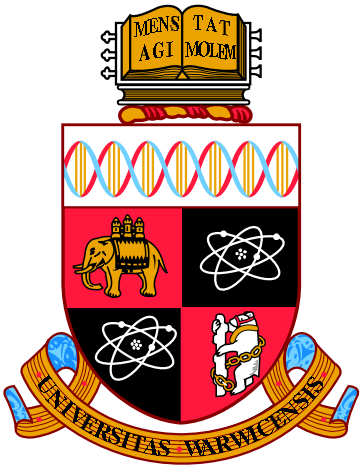
.....

.....

.....

.....

.....



Markov-functional and Stochastic Volatility modelling

by

Duy Pham

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Department of Statistics

September 2012

THE UNIVERSITY OF WARWICK

Contents

List of Tables

List of Figures

Acknowledgments

First and foremost, I am indebted to my supervisor, Dr Jo Kennedy, for her guidance and continuous support. Jo, it wouldn't have been possible without you. Thank you for being incredibly patient with me.

I am grateful to the department of statistics for the financial support I have received and for all the excellent lectures I have learnt from. This has been an incredible experience for me.

I would like to thank Subhankar Mitra for introducing me to the world of coding while we were working together in summer 2010. Thank you for sharing with me your invaluable experience. I am also grateful to a number of anonymous referees who provided valuable comments and suggestions to earlier versions of the papers that form the basis for this thesis.

One paragraph is surely not enough for my friends but I will try my best. First of all, a special thank to all my office mates. I can't imagine how my office life would have been without them. Hasin and Fiona, thank you for all the joy we had together. Giorgos and Javier, those Friday glasses of wine will stay in my memory for a long time. Tomek, you are a true philosopher. Helen, 12pm Friday is always in my diary. For those who were not mentioned, I want to thank you all: from Warwick and off-Warwick, Statistics and non-Statistics. You guys have taught me so many lessons: both academic and non-academic. Thank you all for being there when it matters and spending such a priceless time at Warwick with me.

Finally, I'd like to thank my parents, my grandma, and my sister for their enduring love and support. Thank you for loving me all these years.

Declarations

I declare that this thesis is based on my own research in accordance with the regulations of the University of Warwick. The work is original except where indicated by specific references in the text. Chapter 3 is based on the working paper Kennedy and Pham [2011] (which was submitted for publication). Chapter 5 is based on Kennedy et al. [Forthcoming] (which was accepted for a publication in the journal of Applied Mathematical Finance). The thesis has not been submitted for examination at any other university.

Abstract

In this thesis, we study two practical problems in applied mathematical finance. The first topic discusses the issue of pricing and hedging Bermudan swaptions within a one factor Markov-functional model. We focus on the implications for hedging of the choice of instantaneous volatility for the one-dimensional driving Markov process of the model. We find that there is a strong evidence in favour of what we term "parametrization by time" as opposed to "parametrization by expiry". We further propose a new parametrization by time for the driving process which takes as inputs into the model the market correlations of relevant swap rates. We show that the new driving process enables a very effective vega-delta hedge with a much more stable gamma profile for the hedging portfolio compared with the existing ones.

The second part of the thesis mainly addresses the topic of pricing European options within the popular stochastic volatility SABR model and its extension with mean reversion. We investigate some efficient approximations for these models to be used in real time. We first derive a probabilistic approximation for three different versions of the SABR model: Normal, Log-Normal and a displaced diffusion version for the general constant elasticity of variance case. Specifically, we focus on capturing the terminal distribution of the underlying process (conditional on the terminal volatility) to arrive at the implied volatilities of the corresponding European options for all strikes and maturities. Our resulting method allows us to work with a variety of parameters which cover long dated options and highly stress market condition. This is a different feature from other current approaches which rely on the assumption of very small total volatility and usually fail for longer than 10 years maturity or large volatility of volatility.

A similar study is done for the extension of the SABR model with mean reversion (SABR-MR). We first compare the SABR model with this extended model in terms of forward volatility to point out the fundamental difference in the dynamics of the two models. This is done through a numerical example of pricing forward start options. We then derive an efficient probabilistic approximation for the SABR-MR model to price European options in a similar fashion to the one for the SABR model. The numerical results are shown to be still satisfactory for a wide range of market conditions.

Chapter 1

Introduction

The purpose of this thesis is to contribute further to the current literature of two distinct areas in applied mathematical finance: Markov-functional and stochastic volatility modelling. Since the two topics have little overlap, we treat them separately in Part I and Part II respectively. For each part, we provide a separate background chapter (Chapters 2 and 4) so this introductory chapter only gives a brief outline and summary of the contents of the whole thesis. Readers can choose to read the background chapter of each part and then return here for a better overview of the thesis.

Part I of the thesis discusses the issue of pricing and hedging Bermudan swaptions within a one factor Markov-functional model. We focus on the implications for hedging of the choice of instantaneous volatility for the one-dimensional driving Markov process of the model. The second part of the thesis mainly addresses the topic of pricing European options within the popular stochastic volatility SABR model and its extension with mean reversion. We investigate some efficient approximations of the models to be used in real time.

The mathematical background assumed in this thesis can be accessed from a number of modern textbooks in mathematical finance and interest rate modelling. For example, Karatzas and Shreve [1991] is a standard textbook for various topics in probability and stochastic calculus, and some applications to finance. Another reference is Hunt and Kennedy [2004] which presents a self-contained discussion on the topics ranging from theoretical issues such as Girsanov's theorem and the Martingale Representation theorem to more practical issues regarding arbitrage-free derivatives pricing in the real world. Topics on interest rate and stochastic volatility modelling have been addressed in various texts in the literature. For example, Rebonato [2002], Rebonato [2004] and Brigo and Mercurio [2001] present many different issues on pricing and hedging interest rate derivatives with and without stochastic volatility, and the complete three volumes of Andersen and Piterbarg [2010] is another excellent reference. Readers are not required to read those texts

to get a grasp of interest rate and stochastic volatility modelling. In this thesis, we provide a self-contained study of these topics within specific contexts and refer to relevant materials in the literature whenever necessary.

In Part I (Chapters 2 and 3), we focus on further development of one factor Markov-functional models which are efficient arbitrage-free interest rate models. Chapter 2 gives relevant background of interest rate markets in terms of both products and existing models. We discuss from basic concepts and the workings of various interest rate products to some current issues in the literature. One of the issues is the problem of hedging Bermudan swaptions within one factor Markov-functional models. The questions that we want to answer are: how can different parametrizations of the model lead to different behaviours of the hedging portfolios of Bermudan swaptions? What are the key factors that affect the price and the hedge quality, and how can we set up a one factor model that can improve the hedge?

Bearing these questions in mind, the main contribution that we want to make is discussed in Chapter 3. For a working paper version, see Kennedy and Pham [2011]. In Chapter 3, we consider different choices of the instantaneous volatility for the one-dimensional driving Markov process of the corresponding swap Markov-functional model. One very popular choice is to take a Gaussian process with exponential instantaneous volatility, referred to as the mean reversion process (MR), as is done in Pietersz and Pelsser [2010]. Our first investigation is to compare this candidate with one based on the Hull-White short-rate model, referred to as the Hull-White process (HW) which was first introduced in Bennett and Kennedy [2005].

For these two driving processes, the vega profiles of a Bermudan swaption under the swap Markov-functional model turn out to have some key differences. These differences can be traced back to the difference in nature of the two parametrizations for the driving process. The MR process is an example of what we term "parametrization by expiry". Here the auto-correlations of the driving process are chosen at the outset and controlled by parameters which are user inputs. As such the changes in the correlations of swap rates at their setting dates relevant to the pricing of a Bermudan swaption are not hedged. In contrast, the HW process is an example of "parametrization by time". In this type of parametrization, the auto-correlations of the driving process are linked explicitly to market implied volatilities and it is this important feature which allows the possibility of hedging against moves in market correlations of relevant swap rates. See Section 2.2.2 in Chapter 2 for an intuitive discussion on these concepts via a simple example of a two-period Bermudan swaption.

Based on the investigation of the MR and HW processes, a new parametrization by time for the driving process is proposed in Section 3.3.3. This new parametriza-

tion takes as inputs into the model the market correlations of relevant swap rates. As far as we are aware of the one factor Markov-functional model literature, Kennedy and Pham [2011] was the first to present this idea. We find that this new parametrization has a vega response spread over the swaption matrix but interestingly the total vega for each expiry (row of the swaption matrix) is approximately the same as for the HW model. In Section 3.4.4, a theoretical proof is given to explain this result.

The different vega profiles of the parametrizations by expiry and by time have a direct consequence for hedging. In Section 3.5, we find that when the driving process is parameterized by time the "total" gamma (sum of all gammas) of a vega-delta neutral portfolio for a Bermudan swaption is stabilized. In contrast, it is not possible to control the "total" gamma for this portfolio with the vega profile associated with parametrization by expiry. We further find that the proposed parametrization by time for the driving process with a vega response spread over the swaption matrix leads to a more stable "parallel" gamma profile (sum of each row of the gamma matrix) than that of the HW process. Apart from the gamma results, we also find a much better gamma-theta balance for the hedging portfolios associated with parametrization by time which directly affects the potential Profit and Loss accounts of the portfolios. These findings support our overall view in favour of the parametrization by time for the one-dimensional driving Markov process of the swap Markov-functional model. In some parts of Chapters 2 and 3, we note that this conclusion should apply to one factor separable market models as well which are also popular interest rate models in the current literature.

In Part II (Chapters 4,5,6 and 7), we turn our attention to the topic of stochastic volatility modelling. By convention, market prices of European options are usually quoted in terms of (Black) implied volatilities which are the volatility parameters in the Black formula (Black and Scholes [1973] and Black [1976]). Volatility smile is the phenomenon in the market that different implied volatilities are quoted for European options that have different strikes. The term "smile" arises since implied volatility as a function of strikes sometimes displays the U-shape or smile shape. Here we use the general term "smile" but note that it covers other shapes as well, e.g. skew as seen in various markets. Chapter 4 gives a brief historic background of different volatility models and describes how smile modelling is tackled in the literature. We also discuss at the end of Chapter 4 our thoughts and objectives that we want to achieve in the later chapters.

In Chapter 5, we study one of the most popular stochastic volatility models in practice: the SABR model that was first proposed in Hagan et al. [2002]. See Kennedy et al. [Forthcoming] for the paper version which was accepted for publication in the Journal of Applied Mathematical Finance. The main technical contribution that we make in this chapter is to derive an efficient approximation

to price European options and obtain implied volatilities within a short amount of time. The main idea is to focus on capturing the terminal distribution of the underlying process (conditional on the terminal volatility) to arrive at the implied volatilities of the corresponding European options for all strikes and maturities. The resulting method allows us to work with a variety of parameters which cover long dated options and highly stress market condition. This is a different feature from other current approaches which rely on the assumption of very small total volatility and usually fail for maturity longer than 10 years or large volatility of volatility. We numerically compare this approximation with others in the literature and find that ours outperforms them in most market scenarios.

Chapter 6 is similar to Chapter 5 in terms of objectives but the underlying model that we work with is an extension of the SABR model with mean-reverting volatility (denoted by SABR-MR in this thesis). We explain why practitioners might want to use the SABR-MR model in various contexts, e.g. equity or fixed income. The difference in dynamics between the SABR and SABR-MR models is also addressed via some numerical examples involving pricing forward start options. Through this investigation, we want to make a point that while the two models may be quite similar when pricing European options, there are still some fundamental differences if we are going to use them to price other products.

Later in Chapter 6, we derive an efficient approximation for the SABR-MR model using a similar method to that used in Chapter 5. The results show that the method continues to work well even though the underlying model is more complicated. Various numerical tests for the approximation are performed and compared against Monte Carlo solution, and all yield satisfactory results.

Chapter 7 examines the hedging properties of both the SABR and SABR-MR models. The primary interest is to assess whether the two models yield different hedging results even though they can give similar prices for the same European option. The results show that they still continue to be qualitatively similar from the hedging perspective. As we discuss hedging within stochastic volatility models, we review hedging methodologies from different points of view (theory and practice). This is a topic that has not been dealt with very well in the literature. We do not give an answer to how practitioners can obtain the best hedge with stochastic volatility but we attempt to connect different ideas of hedging European options.

Part I

Markov-functional modelling

Chapter 2

An overview of interest rate modelling

The purpose of this chapter is to give a brief overview of interest rate modelling and set up the foundation for Chapter 3. This chapter will be divided into 2 main parts: products and models. We first describe the main interest rate products and general market practice that will be relevant to the thesis. The models will then be emphasized in terms of their usage for our main application. At the end of the chapter, we provide a simple example to illustrate our intuition for the main work in the next chapter.

2.1 Basic products and derivatives in the interest rate market

In this section, we give a brief outline of the interest rate market in terms of the products and how they work in practice. We start from the most fundamental instruments in the market that are relevant to the thesis. These products include: pure discount bonds, deposits, forward rate agreements and interest rate swaps. We then proceed to different types of options including both vanilla and exotic based on these fundamental instruments. The particular interest rate derivatives that we consider are caplets, European swaptions, and Bermudan swaptions. The materials of this section are gathered in a number of modern text books, e.g. Hunt and Kennedy [2004], Pelsser [2000] and Andersen and Piterbarg [2010]. Here, we set up our own notations that will be consistent throughout Part I of the thesis.

2.1.1 Fundamental Instruments

Amongst the most fundamental products in interest rate markets is the family of *zero-coupon bonds*, also known as the pure discount bonds. Assume that a guaran-

teed payment of a unit currency at time T is a traded product at any time $t < T$ (cashflows illustrated in the figure below). The time-t value of this product is the price of a pure discount bond maturing at T as seen at t. We denote it by D_{tT} for $0 \leq t \leq T$. It is clear that $D_{tt} = 1$ for all $t > 0$. The construction of the family of pure discount bonds for a discrete collection of maturities is commonly referred to as the construction of the yield curve or interest rate curve.

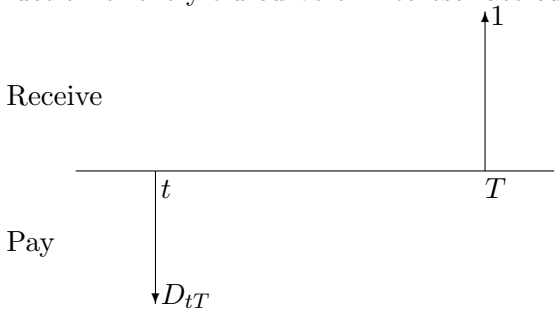


Figure 2.1: Zero-coupon bond

A deposit is an agreement between two counterparties in which one pays the other a cash amount and in return receives this money back at some pre-agreed future date, with a pre-agreed additional payment of interest (proportional to the amount of cash initially deposited).

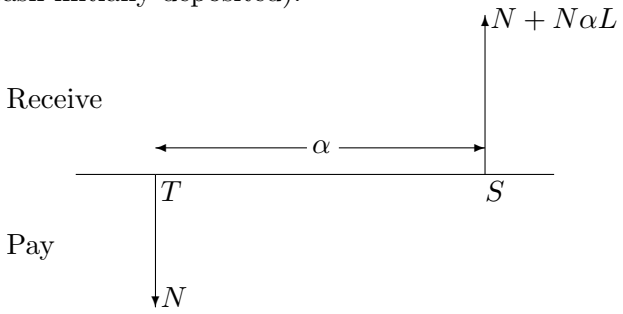


Figure 2.2: Deposit cashflows

In figure 2.2 above, N is the amount of cash deposited at time T and L is the pre-agreed interest rate for the period $[T, S]$. The duration α between T and S is referred to as the accrual factor. Deposits are available for a range of maturities and only a small number of them are quoted as standard.

In the interbank market, L is often known as the (spot) LIBOR rate whose value is known at time T. LIBOR stands for *London Interbank Offered Rate* and it is the rate of interest that one London bank will offer to pay on a deposit by another. In general, LIBOR rates are quoted on values of the accrual factor α ranging from one week to 12 months. The spot LIBOR for the period $[T, S]$ will be denoted by $L_T[T, S]$. One can easily see the relation between the LIBOR rate and discount bonds via figure 2.2. If we deposit the amount of cash D_{TS} in the bank, at time S

we will receive

$$\begin{aligned} 1 &= D_{TS}(1+\alpha L_T[T,S]) \\ \iff L_T[T,S] &= \frac{1-D_{TS}}{\alpha D_{TS}}. \end{aligned} \tag{2.1}$$

While a deposit locks in an interest rate for a given period of time starting immediately, many market participants prefer to lock in interest rates for a given period of time starting in the future. In the fixed income market, contracts that give such an agreement are known as forward rate agreements. A forward rate agreement (FRA) is an agreement between two counterparties to exchange cash payments at some specified date in the future. An FRA involves a notional amount N and two dates, the reset date T and the payment date $S > T$. These two dates will be such that the period $[T, S]$ corresponds exactly to the accrual period of a standard deposit starting at date T. Under an FRA, the first counterparty pays to the second an amount $N\alpha K$, where the accrual factor α is defined as previously and K is the fixed rate that is known when the agreement is made. In return, the second counterparty pays to the first $N\alpha L_T[T, S]$, where $L_T[T, S]$ is the spot LIBOR that sets at T (not known when the agreement is made).

Figure 2.3: FRA cashflows

Assume that two counterparties enter an FRA at time t with no additional cost for some fixed rate K. It is this value of K that is quoted in the FRA market and it is referred to as the forward LIBOR rate, denoted by $L_t[T, S]$ for $t \in [0, T]$. Following a simple static replicating portfolio argument, we obtain the time-t value of an FRA which is equal to $N[D_{tT} - (1 + \alpha K)D_{tS}]$. The forward LIBOR $L_t[T, S]$ can then be expressed in terms of the discount bonds as follows

$$L_t[T,S] = \frac{D_{tT} - D_{tS}}{\alpha D_{tS}}. \tag{2.2}$$

An interest rate swap, or swap, is an agreement between two counterparties to exchange a series of cashflows on pre-agreed dates in the future. Let $0 = T_0 < T_1 < \cdots < T_{n+1}$ denote a given tenor structure and the accrual factors $\alpha_i = T_{i+1} - T_i$ for $i = 1, \ldots, n$. We denote the forward LIBOR $L_t[T_i, T_{i+1}]$ by L_t^i for brevity. The payment of a swap may be decomposed into the floating legs $N\alpha_i L_{T_i}^i$ and the fixed legs $N\alpha_i K$ at T_{i+1} where K is the fixed rate for the swap. The contract with fixed

pays is referred to as a payers swap whereas the one with reversed cashflows is a receivers swap. Note that a payers swap could be valued in exactly the same way as for an FRA. Hence, one can easily see that the value of the above swap at time $t < T_1$ is given by

$$V_{\text{swap}}^{1,n}(t) = N(D_{tT_1} - D_{tT_{n+1}} - K\sum_{k=1}^{n} \alpha_k D_{tT_{k+1}}).$$

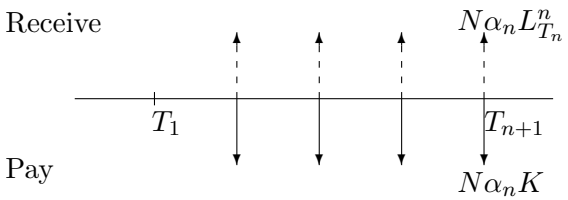


Figure 2.4: Payers interest rate swap cashflows

In general, one can specify a swap for any given tenor structure. For instance, assume a tenor structure $T_i < T_{i+1} < \cdots < T_{j+1}$, we then obtain exactly the same valuation formula as the above with appropriate indices. The forward swap rate, that we choose to denote by $y_t^{i,j+1-i}$ for $t \in [0, T_i]$, is the fixed rate K for which the time-t value of the corresponding swap is zero. The superscript indicates that the swap with tenor (swap length or number of payments) $j + 1 - i$ is entered at expiry time T_i, i.e. the last payment is made at maturity time T_{j+1}. Similar to the LIBOR rates, swap rates are also the *reference interest rates* that are set by a financial authority. The swap rate and discount bonds can be easily linked by the following relation

$$y_t^{i,j+1-i} = \frac{D_{tT_i} - D_{tT_{j+1}}}{\sum_{k=i}^{j} \alpha_k D_{tT_{k+1}}}. \tag{2.3}$$

The term in the denominator is referred to as the present value of a basis point (PVBP) and is denoted by

$$P_t^{i,j+1-i} = \sum_{k=i}^{j} \alpha_k D_{tT_{k+1}}.$$

Given the swap rate $y^{i,j+1-i}$, we obtain a more convenient formula for the swap

$$V_{\text{swap}}^{i,j+1-i}(t) = NP_t^{i,j+1-i}(y_t^{i,j+1-i} - K).$$

2.1.2 Vanilla interest rate options

In order to protect investors from the possibility that interest rates may increase or decrease in the future, options written on those rates are issued and liquidly traded

in the market. Practitioners usually refer to these options as vanilla options. In what follows, we define some vanilla interest rate options that are most relevant to this thesis.

A caplet is an option on an FRA which pays $V^i_{\text{cl}}(T_{i+1}) := N\alpha_i \max(L^i_{T_i} - K, 0)$ at time T_{i+1} where K is the strike, L^i is the forward LIBOR, and N is the notional. The time-T_i value of this caplet can be discounted from its time-T_{i+1} value, i.e. $V^i_{\text{cl}}(T_i) = N\alpha_i D_{T_iT_{i+1}} \max(L^i_{T_i} - K, 0)$. A floorlet has the opposite payout function to a caplet: $V^i_{\text{fl}}(T_{i+1}) := N\alpha_i \max(K - L^i_{T_i}, 0)$. One can think of a caplet and a floorlet as a call and a put on the LIBOR respectively.

Just as a caplet is an option on an FRA, a European swaption is an option on a swap. The time-t value of the European swaption with the underlying swap's time-t value $V^{i,j+1-i}_{\text{swap}}(t)$ and the associated swap rate $y^{i,j+1-i}_t$ is given by the payoff at time T_i

$$V^{i,j+1-i}_{\text{sption}}(T_i) := \max(V^{i,j+1-i}_{\text{swap}}(T_i), 0). \tag{2.4}$$

In case of a payer swaption with some strike K, its T_i-value can be written in a more convenient form

$$V^{i,j+1-i}_{\text{sption}}(T_i) = NP^{i,j+1-i}_{T_i} \max(y^{i,j+1-i}_{T_i} - K, 0).$$

Similar to caplets, the above form clearly indicates a payer (or receiver) swaption is a call (or put) option on the swap rate. Hence, caplets/floorlets and European swaptions can be priced by a vanilla model[1] which only considers one single rate at a time. We take the swap rate $y^{i,j+1-i}$ as an example. Suppose we work with the martingale measure that takes its PVBP $P^{i,j+1-i}$ as the numeraire. We refer to this measure as the swaption measure (forward measure for caplets and floorlets). The definition in (2.3) implies that $y^{i,j+1-i}$ is a martingale and can be modelled by a driftless Log-Normal process (Black model) under its swaption measure

$$\mathrm{d}y^{i,j+1-i}_t = \sigma_{i,j+1-i} y^{i,j+1-i}_t \mathrm{d}W_t, \qquad \sigma_{i,j+1-i} > 0.$$

In order to evaluate $V^{i,j+1-i}_{\text{sption}}(0)$, one needs to consider the following expectation under the swaption measure

$$V^{i,j+1-i}_{\text{sption}}(0) = NP^{i,j+1-i}_0 \mathbb{E}\left[\frac{P^{i,j+1-i}_{T_i}(y^{i,j+1-i}_{T_i} - K)^+}{P^{i,j+1-i}_{T_i}}\right] = NP^{i,j+1-i}_0 \mathbb{E}[(y^{i,j+1-i}_{T_i} - K)^+].$$

If we carry out full calculation of this expectation, we will then arrive at the famous

[1] We adopt the term *vanilla model* from Andersen and Piterbarg [2010].

Black formula in Black [1976]

$$V_{\text{sption}}^{i,j+1-i}(0) = NP_0^{i,j+1-i}(y_0^{i,j+1-i}\Phi(d_1) - K\Phi(d_2)), \tag{2.5}$$

where $\Phi(\cdot)$ denotes the Normal cumulative distribution function and

$$\begin{aligned} d_1 &= \frac{\ln(y_0^{i,j+1-i}/K)}{\sigma_{i,j+1-i}\sqrt{T_i}} + \frac{1}{2}\sigma_{i,j+1-i}\sqrt{T_i}, \\ d_2 &= d_1 - \sigma_{i,j+1-i}\sqrt{T_i}. \end{aligned}$$

Conversely, market prices of European swaptions and caplets can be translated into the market implied distributions of forward swap rates and LIBOR rates. It is market convention that the market price of the European swaption associated with $y^{i,j+1-i}$ is usually quoted in terms of the Black implied volatility, denoted by $\widetilde{\sigma}_{i,j+1-i}$ for each strike K. That is the value for the volatility parameter that we plug into the Black formula to recover this market price[2]. This means that if we set the volatility parameter $\sigma_{i,j+1-i}$ in the above SDE to be the same as the Black implied volatility $\widetilde{\sigma}_{i,j+1-i}$ for each strike K, the model price computed by the Black formula and the market price of the corresponding European swaption will coincide. One also has a choice to use a deterministic function of time $\sigma_{i,j+1-i}(\cdot)$ instead of the constant $\sigma_{i,j+1-i}$ in the SDE for the swap rate, but has to ensure that $\int_0^{T_i} \sigma_{i,j+1-i}^2(t)\mathrm{d}t = \widetilde{\sigma}_{i,j+1-i}^2 T_i$ for the perfect recovery of market price.

It is observed from the market that for different strikes K, we have different implied volatilities and this phenomenon is known as "smile" (or skew). This problem forms the second part of the thesis where we study various types of vanilla models rather than the Black model. In Part I, we will work without the presence of smile or equivalently assume flat volatility smile curves.

2.1.3 Bermudan swaptions

We have discussed some of the most basic and fundamental products which form important building blocks for other more sophisticated interest rate derivatives. Another popular type of interest rate derivatives is the Bermudan style interest rate derivatives such as Bermudan swaptions. Unlike the previously discussed vanilla options, Bermudan swaptions are not liquidly traded and considered exotic products.

A financial product that has multiple exercise dates is called Bermudan. A Bermudan swaption is an option to enter into a swap on any of its pre-specified fixing dates. We consider a particular example of a co-terminal Bermudan swaption

[2]The market price given by the above Black formula implies the Log-Normal distribution of the swap rate at its expiry date $y_{T_i}^{i,j+1-i}$ under its swaption measure. In practice, they can be given by the Bachelier formula too which implies the corresponding Normal distribution (see Chapter 4).

where all the underlying swaps end at the same date T_{n+1}. To be more specific, the holder of a (co-terminal) Bermudan swaption has the right, on any of the swap exercise dates to enter the remaining swap, i.e. at time T_i, $i = 1, \ldots, n$ enter the swap that ends at T_{n+1}. At each exercise date, the holder maximizes the value of entering the swap now and the value of the Bermudan swaption with the remaining exercise dates (exercise the option later). Note that the Bermudan swaption holder can only exercise the option once. Due to its nature, one can think of a Bermudan swaption as an option on an option (on an option, etc.). Other types of Bermudan swaption include the co-sliding or constant maturity Bermudan swaptions where all the underlying swaps have the same tenor (swap length), or the zero-coupon Bermudan swaptions where the notionals of the underlying swaps are different. See volume 3 of Andersen and Piterbarg [2010] for a detailed discussion on the general class of callable LIBOR exotics which includes Bermudan swaptions.

2.2 Pricing and hedging Bermudan swaptions in practice

While Bermudan swaptions seem to be an attractive product due to the potential profits that investors are entitled to, the complex structure of the product and the source of exotic risks that they contain could be a tremendous problem. One then certainly needs a sophisticated enough model as a pricing and hedging tool for this kind of products.

An obvious difficulty is that the price a Bermudan swaption has to be consistent with the market prices of its underlying European swaptions (known via implied volatilities). Otherwise, the option holder can make arbitrage opportunities out of the inconsistencies in prices. The first criterion that we need is *calibration* where the chosen model can reproduce vanilla prices of the underlying Europeans by using implied volatilities before attempting to quote a Bermudan price. Furthermore, the Bermudan optionality nature of the product indicates that its price depends vastly on multiple rates. As a result, we need to consider the joint distribution of different underlying rates across different time points rather than the marginal distribution of just one single rate at one time. The evaluation process, hence, requires a detailed knowledge of the dynamics of the whole term structure.

Throughout the life of a Bermudan swaption, the rates and market quotes for implied volatilities both change. The model then needs to be re-calibrated to more recent market quotes to update the Bermudan swaption price. Consequently, we face two main risks: delta and vega risks. In this context, delta risk is the risk with respect to the change in the interest rate curve while vega risk is with respect to

the change in implied volatilities[3]. These are the typical examples of in-model and out-of-model hedging concepts. For a comprehensive discussion on this topic, see for example Rebonato [2004]. The practice of hedging requires practitioners to acquire appropriate proportions of other assets/hedging instruments to offset the vega and delta risks. This means that the corresponding portfolio containing the Bermudan swaption and its hedging instruments will not be exposed to any small changes in the interest rate curve and implied volatilities over a short period of time. The risk management of Bermudan swaptions is vital and considered to be as important as the pricing especially when there are many underlyings involved in the product.

Having stressed the importance of hedging, we want to note that this matter has to be considered with care. Different models imply different delta and vega risks. The underlying reason is that even when different models can calibrate to the same market data, it is possible that they can give different hedges for the Bermudan since other factors affect the product as well, e.g. the joint distribution of multiple rates. If we hedge the Bermudan swaption assuming wrong delta and vega risks, the hedging cost could be substantially high and lead to potentially big losses for investors. Bearing this in mind, choosing a good model for the pricing and hedging purposes is another important issue for practitioners.

2.2.1 The choice of models

One of the main challenges in the development of interest rate models is the calibration property. Although models in the past such as short-rate models are highly tractable and relatively simple to understand, e.g. Hull and White [1990], it is difficult to calibrate them to the interest rate curve and the set of relevant vanilla prices for pricing exotic products. The reason is that the hypothetical short rate is not directly observable in the market. Obviously, in terms of pricing a Bermudan swaption there is no straightforward way of using implied volatilities of European swaptions as input to the short-rate models. See Chapter 17 of Hunt and Kennedy [2004] for an example of using the Vasicek-Hull-White model to price a Bermudan swaption.

In the late 90s, the emergence of market models marked a breakthrough in the interest rate modelling literature. The class of market models include both LIBOR and swap market models, also known as BGM models (see Brace et al. [1997], Miltersen et al. [1997] and Jamshidian [1997]). The big advantage of these models is the calibration property where practitioners assume that the rates that are directly observable from the market (forward LIBORs or forward swap rates) are Log-Normal martingales in their own (forward or swaption) measures. Market models, therefore, allow for a straightforward calibration to market vanilla prices

[3]See Chapter 3 and Chapter 7 for different forms of delta and vega risks

using the quoted Black implied volatilities in the Black formula for caplets and European swaptions. See Rebonato [2002] and Rebonato [2004] for a comprehensive review of the development of market models for pricing and hedging interest rate derivatives.

Another challenge in interest rate modelling is the efficiency in implementation as practitioners have to produce prices within a reasonably short period of time. Despite the calibration advantage and clear intuition, market models suffer from their high dimensionality[4] even when we use only one factor for the SDEs of all LIBORs or swap rates. In terms of pricing callable products like Bermudan swaptions in market models, Monte Carlo simulation is needed to determine the exercise boundary. Different approximations have been developed in the past in order to reduce the computational burden of Monte Carlo simulation, e.g. see Longstaff and Schwartz algorithm in Longstaff and Schwartz [2001]. In the literature, the common way to get around the high dimensionality problem is by approximating the market model by some process of lower dimension (see Pietersz et al. [2004]) under an extra assumption of *separability*. Unfortunately, this method in general encourages a significant amount of arbitrage in the model.

The class of Markov-functional models has been introduced by Hunt et al. [2000] to overcome the shortcomings of market models (see also Chapter 19, Hunt and Kennedy [2004] and Pelsser [2000]). Markov-functional models can fit the observed prices of vanilla instruments similarly to market models (consistent with the Black formula), but at the same time they inherit the practical advantage from short-rate models as we only need to keep track of some low (usually one or two) dimensional Markov process. Markov-functional models are typically implemented on a lattice under the terminal measure (taking the terminal discount bond as numeraire) but other versions of Markov-functional models exist too. For example, the cross-currency and hybrid Markov-functional models were presented in Fries and Rott [2004] and Fries and Eckstaedt [2009] under the spot measure (taking the money market account as numeraire). See also Fries [2007] for tips on the implementation of these models. Another recent development on Markov-functional models is the n-dimensional LIBOR Markov-functional model introduced in Kaisajuntti and Kennedy [Forthcoming] also under the spot measure. The implementation of this model requires Monte Carlo simulation rather than the typical lattice implementation as for the one-dimensional version.

The one-dimensional (or one factor[5]) Markov-functional model also has some possible drawbacks. In terms of pricing and hedging Bermudan swaptions, one

[4]The dimension of the model grows with the number of rates being modelled. In some modern texts, some author refer to this as "the curse of dimensionality".

[5]Unlike market models, the number of factors used for the driving Markov process is equal to the number of dimensions of the corresponding Markov-functional model.

disadvantage is that we have only one Brownian motion to control the correlation structure of the model that could have a certain effect on the price and hedge. Consequently, the instantaneous correlation of different rates could only be one. In the literature, one of the biggest debates is whether it is necessary to use a multi-factor model to enable a more flexible correlation structure for pricing and hedging Bermudan swaptions. A good summary of the current literature on this topic is given in Pietersz and Pelsser [2010] or Chapter 6 of Pietersz [2005]. We also list here a number of references that compared single factor models with multi-factor models with different views and opinions: Longstaff et al. [2001], Driessen et al. [2003], and Fan et al. [2003]. Back to the paper by Pietersz and Pelsser [2010], the authors carry out a comparison of the hedging performance of a single factor Markov-functional model and multi-factor market models in relation to Bermudan swaptions and their findings support the claim that if a one factor Markov-functional model is appropriately calibrated to "terminal correlations"[6] of swap rates that are relevant to the Bermudan swaption then the hedging performance of both the multi-factor and one factor models are comparable. See also Pietersz [2005] for more technical discussions on using multi-factor market models for Bermudan-style interest rate derivatives.

In spirit of the work by Pietersz and Pelsser [2010], we want to initiate our own investigation on the pricing and hedging of Bermudan swaptions within a one factor Markov-functional model driven by a Gaussian process. The main contribution we want to make is to study the implications for hedging of the choice of the instantaneous volatility for the driving process. This is a topic which have received little attention in the literature for one factor Markov-functional models or equivalently for one factor market models under the assumption of separability (see Bennett and Kennedy [2005] and Pietersz et al. [2004]). We will discuss this in Chapter 3. We conclude the current chapter by including a worked-out example in the next subsection to illustrate our intuition.

2.2.2 A motivational example: two-period Bermudan swaption

Let us consider a concrete example of a two-period Bermudan swaption to get a flavour of the problem that we are going to tackle in Chapter 3. This simple two-period Bermudan swaption involves three different rates: the one-year LIBOR L^1, the two-year LIBOR L^2, and the one-year into two-year swap rate $y^{1,2}$. The underlying swaption values for this Bermudan swaption are $V^{1,2}_{\text{sption}}(0)$ and $V^{2,1}_{\text{sption}}(0)$ with the market Black implied volatilities $\widetilde{\sigma}_{1,2}$ and $\widetilde{\sigma}_{2,1}$ respectively. Note that the two-year LIBOR L^2 coincides with the two-year into one-year swap rate $y^{2,1}$, but

[6]The term "terminal correlation" has slightly different interpretations in the literature. When we use this term later in the thesis, we will specify exactly what we mean.

we will use the former rather than the latter to avoid confusion.

Suppose we want to use a one factor LIBOR market model to price this product. The model assumes the following dynamics for the two LIBOR rates under the same (unspecified) measure

$$\mathrm{d}L_t^i = \mu_t^i L_t^i \mathrm{d}t + \sigma_i(t) L_t^i \mathrm{d}W_t, \qquad i = 1, 2. \tag{2.6}$$

Recall that each LIBOR is a Log-Normal martingale under its own forward measure, i.e. no drift in the SDE. We ignore the drift term in (2.6) for now for some reasons to be discussed in Chapter 3, so the involvement of measure is neglected here and only the diffusion part is assumed to be significant. The only flexibility we have here is the specification of the instantaneous volatility functions $\sigma_i(t)$ which will have a direct link to the Black formula and Black implied volatilities in (2.5). We consider two different parametrizations for the instantaneous volatility functions $\sigma_i(t)$ for $0 \leq t \leq T_i$ and $i = 1, 2$.

1. *Parametrization by expiry* (constant volatility): $\sigma_i(t)$ is a positive constant throughout its life, i.e. $\sigma_i(t) = \sigma_i$ for $0 \leq t \leq T_i$.

2. *Parametrization by time* (piecewise-constant volatility): $\sigma_i(t)$ is chosen to be piecewise-constant, i.e.

$$\begin{aligned} \sigma_1(t) &= \sigma_1, \qquad t \in [0, T_1] \\ \sigma_2(t) &= \sigma_{2,1}, \qquad t \in [0, T_1] \\ &= \sigma_{2,2}, \qquad t \in (T_1, T_2]. \end{aligned}$$

   All the constants here must be positive. We assume the two log-LIBORs have the same instantaneous volatility for the first period, i.e. $\sigma_{2,1} = \sigma_1$ for $0 \leq t \leq T_1$. Note that for the previous parametrization by expiry case, we basically impose that $\sigma_{2,1} = \sigma_{2,2} = \sigma_2$ equivalently.

We now require that the two underlying swaptions should be simultaneously correctly evaluated. This is the compulsory model calibration procedure. It is clear for the two-year caplet that we need to ensure the following link between the instantaneous volatility function of the log-LIBOR and the market Black implied volatility

$$\widetilde{\sigma}_{2,1}^2 T_2 = \int_0^{T_2} \sigma_2^2(t) \mathrm{d}t. \tag{2.7}$$

For the one-year into two-year swaption, the issue is a bit more delicate since we do not have a direct control over the instantaneous volatility of the log of the corresponding swap rate to link to its implied volatility. However, we can use the connection of this swap rate with the two underlying LIBORs and some further

approximations to arrive at the following condition

$$\widetilde{\sigma}_{1,2}^2 T_1 \approx \int_0^{T_1} (\zeta_1 \sigma_1(t) + \zeta_2 \sigma_2(t))^2 \mathrm{d}t, \tag{2.8}$$

where ζ_1 and ζ_2 are assumed to be positive constant weights (see Chapter 3 or Rebonato [2002] for their full expressions and explanations). This is because the instantaneous volatility $\sigma_{1,2}(t)$ of the log of the swap rate $y_t^{1,2}$ is approximately $\zeta_1\sigma_1(t) + \zeta_2\sigma_2(t)$ in a one factor model (see Chapter 3), i.e.

$$\mathrm{d}y_t^{1,2} \approx \ldots \mathrm{d}t + (\zeta_1 \sigma_1(t) + \zeta_2 \sigma_2(t)) y_t^{1,2} \mathrm{d}W_t. \tag{2.9}$$

For the parametrization by expiry, no choice is left for σ_2 as the condition from (2.7) uniquely determines σ_2

$$\begin{aligned} \widetilde{\sigma}_{2,1}^2 T_2 &= \sigma_2^2 T_2 \\ \iff \sigma_2 &= \widetilde{\sigma}_{2,1}. \end{aligned} \tag{2.10}$$

Now, by using (2.8) σ_1 can be solved explicitly as follows

$$\begin{aligned} \widetilde{\sigma}_{1,2}^2 T_1 &\approx (\zeta_1\sigma_1 + \zeta_2\sigma_2)^2 T_1 \\ \iff \widetilde{\sigma}_{1,2} &\approx \zeta_1\sigma_1 + \zeta_2\widetilde{\sigma}_{2,1} \\ \iff \sigma_1 &\approx \frac{\widetilde{\sigma}_{1,2} - \zeta_2\widetilde{\sigma}_{2,1}}{\zeta_1}. \end{aligned} \tag{2.11}$$

For the parametrization by time, the condition (2.8) implies that

$$\begin{aligned} \widetilde{\sigma}_{1,2}^2 T_1 &\approx (\zeta_1\sigma_1 + \zeta_2\sigma_{2,1})^2 T_1 \\ \iff \widetilde{\sigma}_{1,2}^2 &\approx \sigma_1^2(\zeta_1 + \zeta_2)^2 \\ \iff \sigma_1 &\approx \frac{\widetilde{\sigma}_{1,2}}{\zeta_1 + \zeta_2}. \end{aligned} \tag{2.12}$$

We recall that the second line of the above equations follows because $\sigma_{2,1} = \sigma_1$ by the initial assumption. From the derived expression for σ_1, we then find the expression for $\sigma_{2,2}$ by using (2.7)

$$\begin{aligned} \widetilde{\sigma}_{2,1}^2 T_2 &= \int_0^{T_1} \sigma_{2,1}^2 \mathrm{d}t + \int_{T_1}^{T_2} \sigma_{2,2}^2 \mathrm{d}t \\ &= \sigma_1^2 T_1 + \sigma_{2,2}^2 (T_2 - T_1) \\ \iff \sigma_{2,2} &\approx \sqrt{\frac{\widetilde{\sigma}_{2,1}^2 T_2 - \frac{\widetilde{\sigma}_{1,2}^2}{(\zeta_1+\zeta_2)^2} T_1}{T_2 - T_1}}. \end{aligned} \tag{2.13}$$

For this two-period Bermudan swaption, we want to note that the future realizations of $y^{1,2}_{T_1}$ or $L^2_{T_2}$ determine the future payoffs of the product if one is going to exercise the option at T_1 or T_2 respectively. In practice, the option holder can choose to exercise at T_1 if the payoff is positive. Nevertheless, even when the immediate exercise value is positive, the holder can decide to hold on to the swaption in view of a more favourable $L^2_{T_2}$. In that sense, it is clear that the today's value of this two-period Bermudan swaption should be influenced heavily by the joint distribution of $y^{1,2}_{T_1}$ and $L^2_{T_2}$ as well. It was suggested in a number of references, e.g. Rebonato [2002] and Brigo and Mercurio [2001], that the correlation $\mathrm{Corr}(\ln y^{1,2}_{T_1}, \ln L^2_{T_2})$ is indeed one of the main driving forces of the price. Here we consider this correlation term explicitly

$$\begin{aligned}
\mathrm{Corr}(\ln y^{1,2}_{T_1}, \ln L^2_{T_2}) &= \frac{\mathrm{Cov}(\ln y^{1,2}_{T_1}, \ln L^2_{T_2})}{\sqrt{\mathrm{Var}(\ln y^{1,2}_{T_1})}\sqrt{\mathrm{Var}(\ln L^2_{T_2})}} \\
&\approx \frac{\mathrm{Cov}(\ln y^{1,2}_{T_1}, \ln L^2_{T_1})}{\widetilde{\sigma}_{1,2}\sqrt{T_1}\widetilde{\sigma}_{2,1}\sqrt{T_2}} \\
&\approx \frac{\int_0^{T_1}(\zeta_1\sigma_1(t)+\zeta_2\sigma_2(t))\sigma_2(t)\mathrm{d}t}{\widetilde{\sigma}_{1,2}\sqrt{T_1}\widetilde{\sigma}_{2,1}\sqrt{T_2}} \\
&\approx \frac{(\zeta_1\sigma_1+\zeta_2\sigma_{2,1})\sigma_{2,1}T_1}{\widetilde{\sigma}_{1,2}\sqrt{T_1}\widetilde{\sigma}_{2,1}\sqrt{T_2}}.
\end{aligned}$$

Recall that we did not specify the measure and assumed earlier that only the diffusion parts of the LIBORs and swap rates matter. The numerator in the second line of the above equations follows immediately from the independence of increment in (2.6) and (2.9), while the denominator follows from the implication by the Black formula.

For the parametrization by expiry, recall that $\sigma_{2,1} = \sigma_{2,2} \neq \sigma_1$ and from (2.10) and (2.11) we obtain the following

$$\mathrm{Corr}(\ln y^{1,2}_{T_1}, \ln L^2_{T_2}) \approx \frac{\left(\zeta_1\left(\frac{\widetilde{\sigma}_{1,2}-\zeta_2\widetilde{\sigma}_{2,1}}{\zeta_1}\right)+\zeta_2\widetilde{\sigma}_{2,1}\right)\widetilde{\sigma}_{2,1}\sqrt{T_1}}{\widetilde{\sigma}_{1,2}\widetilde{\sigma}_{2,1}\sqrt{T_2}} = \sqrt{\frac{T_1}{T_2}}. \tag{2.14}$$

For the parametrization by time, recall that $\sigma_1 = \sigma_{2,1} \neq \sigma_{2,2}$ and from (2.12) and (2.13) we have that

$$\mathrm{Corr}(\ln y^{1,2}_{T_1}, \ln L^2_{T_2}) \approx \frac{(\zeta_1+\zeta_2)\frac{\widetilde{\sigma}^2_{1,2}}{(\zeta_1+\zeta_2)^2}\sqrt{T_1}}{\widetilde{\sigma}_{1,2}\widetilde{\sigma}_{2,1}\sqrt{T_2}} = \frac{\widetilde{\sigma}_{1,2}}{\widetilde{\sigma}_{2,1}(\zeta_1+\zeta_2)}\sqrt{\frac{T_1}{T_2}}. \tag{2.15}$$

It is clear that the expressions in (2.14) and (2.15) are quite different mathematically. While the formula in (2.15) takes into account the market Black implied

volatilities, the other does not. This difference leads to the two following important questions regarding the hedging issue:

1. How will the correlation structure of the rates alter as the implied volatilities change?

2. Consequently, how will the price of this two-period Bermudan swaption be affected as a result of the change in correlation structure?

Of course, the two different parametrizations of volatility functions lead to two models that respond very differently to the change in market data. Potentially, the answers to the above questions will also vary substantially between the two models as we shall see in the next chapter. We want to emphasize at this point that this feature is what makes the hedging of Bermudan swaptions somewhat delicate and so crucial in practice. If we parametrize the model in a way that is not realistic and consistent with the market, the hedging error could be devastating and cost the option holder a fortune. We carry out our analysis further in Chapter 3 where we study the problem of hedging Bermudan swaptions within a one factor Markov-functional model specifically.

Chapter 3

Implications for Hedging of the choice of driving process for one-factor Markov-functional models

In this chapter, we study the implications for hedging Bermudan swaptions of the choice of the instantaneous volatility for the driving Markov process of the one-dimensional swap Markov-functional model. We find that there is a strong evidence in favour of what we term "parametrization by time" as opposed to "parametrization by expiry". We further propose a new parametrization by time for the driving process which takes as inputs into the model the market correlations of relevant swap rates. We show that the new driving process enables a very effective vega-delta hedge with a much more stable gamma profile for the hedging portfolio compared with the existing ones.

3.1 Introduction

The problem of pricing and hedging a Bermudan swaption has been of great practical concern in the fixed income quantitative research. The product itself is among the most common exotic interest rate derivatives. However, opinions differ as to what constitutes an effective modelling framework for pricing and hedging Bermudan swaptions. As noted in Chapter 2, one of the biggest debates is whether it is necessary to use a multi-factor model. A good summary of the current literature on this topic is given in Pietersz and Pelsser [2010]. We also recall from Chapter 2 that in Pietersz and Pelsser [2010], the authors carry out a comparison of the hedging performance of a single factor Markov-functional model and multi-factor market

models in relation to Bermudan swaptions. Their findings support the claim that if a single factor Markov-functional model is appropriately calibrated to "terminal correlations" of swap rates that are relevant to the Bermudan swaption then the hedging performance of both the multi-factor and single factor models are comparable.

In this chapter, we restrict attention to the pricing and hedging of Bermudan swaptions within the context of a one factor Markov-functional model driven by a Gaussian process. The contribution we make here is to study the implications for hedging of the choice of the instantaneous volatility for the driving Markov process. This is a topic which seems to have received little attention in the literature for one factor Markov-functional models or equivalently for one factor separable market models (see Bennett and Kennedy [2005] and Pietersz et al. [2004]). One popular choice is to take a Gaussian process with exponential instantaneous volatility, referred to as the mean reversion process (MR), as is done in Pietersz and Pelsser [2010]. We begin our investigation by comparing this candidate with one based on the Hull-White short-rate model, referred to as the Hull-White process (HW) which was first introduced in Bennett and Kennedy [2005].

For these two candidate processes the vega profiles of a Bermudan swaption under the swap Markov-functional model turn out to have some key differences (see also Pertursson [2008] for a comparison of their vega profiles under different market scenarios). These differences can be linked back to the difference in nature of the two parametrizations for the driving process. The mean reversion process (MR) is an example of what we term "parametrization by expiry". Here the auto-correlations of the driving process are chosen at the outset and controlled by parameters which are user inputs. As such the changes in the correlations of swap rates at their setting dates relevant to the pricing of a Bermudan are not hedged. In contrast, the Hull-White process (HW) is an example of "parametrization by time". In this type of parametrization, the auto-correlations of the driving process are linked explicitly to market implied volatilities and it is this feature which allows the possibility of hedging against moves in market correlations of relevant swap rates.

Based on the insight gained by our study of the MR and HW processes, we propose a new parametrization by time for the driving process. This new parametrization takes as inputs into the model the market correlations of relevant swap rates. These market correlations are estimated via a full rank LIBOR market model using a two-step procedure involving a global and local fit to the swaption matrix. This new parametrization has a vega response spread over the swaption matrix but interestingly the total vega for each expiry (row of the swaption matrix) is approximately the same as for the HW model. We give an explanation for why this is the case.

The different vega profiles of the parametrizations by expiry and by time have a direct consequence for hedging. We find that when the driving process is parameterized by time the "total" gamma (sum of all gammas) of a vega-delta neutral portfolio for a Bermudan swaption is stabilized. In contrast, it is not possible to control the "total" gamma for this portfolio with the vega profile associated with parametrization by expiry. We further find that the proposed parametrization by time for the driving process with a vega response spread over the swaption matrix leads to a more stable "parallel" gamma profile (sum of each row of the gamma matrix) than that of the HW process. In addition to the gamma results, we also look at the theta and find a much better gamma-theta balance for the hedging portfolios associated with parametrization by time which affects their potential Profit and Loss accounts to some extent.

The chapter is organized as follows. In Section 3.2, we review the preliminaries and set up the notations. In Section 3.3, we first describe the one-dimensional swap Markov-functional model and analyze the difference between parametrizations by expiry and by time. After that, we construct a new parametrization by time for the driving process. In Section 3.4, we compute the vegas of a Bermudan and analyze them theoretically. A hedging result with an emphasis on the gamma risks will be addressed in Section 3.5. We analyze the gamma-theta balance in Section 3.6. Section 3.7 concludes the chapter.

3.2 Notations and preliminaries

Consider a general tenor structure

$$0 = T_0 < T_1 < \cdots < T_{n+1},$$

where $\alpha_i = T_{i+1} - T_i$ are the accrual factors for $i = 0, \ldots, n$.

Let D_{tT} denote the time-t value of a zero-coupon discount bond that matures at time T. We denote by L^i the forward LIBOR that sets (expires) at T_i and settles (matures) at T_{i+1}. Recall from Chapter 2 that forward LIBORs and discount bonds can be linked via the relation

$$L_t^i = \frac{D_{tT_i} - D_{tT_{i+1}}}{\alpha_i D_{tT_{i+1}}}, \qquad t \leq T_i, \tag{3.1}$$

for $i = 0, \ldots, n$. We denote by $y^{i,j}$ the forward swap rate of an interest rate swap with setting dates $T_i, T_{i+1}, \ldots, T_{i+j-1}$ and settlement dates $T_{i+1}, T_{i+2}, \ldots, T_{i+j}$. Similar to forward LIBORs, forward swap rates can also be written in terms of discount

bonds as noted in Chapter 2

$$y_t^{i,j} = \frac{D_{tT_i} - D_{tT_{i+j}}}{\sum_{k=i}^{i+j-1} \alpha_k D_{tT_{k+1}}}, \qquad t \leq T_i, \tag{3.2}$$

for $i = 0, \ldots, n$. It is clear that $y^{i,1}$ coincides with L^i. For each swap rate $y^{i,j}$, we further introduce the corresponding at the money (ATM) Black implied volatility $\widetilde{\sigma}_{i,j}$.

We recall from Chapter 2 the main features of Bermudan swaptions. The type of Bermudan swaption we consider in this chapter is the co-terminal version, as opposed to other non-standard types of Bermudan swaption. The holder of a co-terminal Bermudan swaption has the right, on any of the swap exercise dates to enter the remaining swap which ends at the pre-determined terminal date T_{n+1}. The underlying swap at T_i consists of a number of coupons that set at T_j and settle at T_{j+1} for $j = i, \ldots, n$. We further denote the notional amount by N and the strike by K. Suppose that the Bermudan swaption is to be exercised at time T_i. In case of a pay fixed type, the holder will then receive the corresponding coupons from the underlying swap, i.e. at T_{j+1} for each $j = i, \ldots, n$ he or she will receive the floating leg $N\alpha_j L^j_{T_j}$ and pay the fixed leg $N\alpha_j K$. In case of a receive fixed type, the holder will receive the fixed legs in exchange for the floating legs. Although the coupons depend on the values of the LIBORs at their setting dates, the exercise value of the underlying swap at each exercise date T_i depends on the corresponding co-terminal swap rate at its setting date $y^{i,n+1-i}_{T_i}$. For a pay fixed Bermudan, the holder will only exercise at time T_i if $y^{i,n+1-i}_{T_i}$ is above the strike level K. Nevertheless, even when the immediate exercise value is positive, the holder can decide to hold on to the swaption in view of a more favourable co-terminal swap rate $y^{j,n+1-j}_{T_i}$ for $j > i$. It was noted in Pietersz and Pelsser [2010] that although the joint distribution of the random variables $\{y^{j,n+1-j}_{T_i}; j = i, \ldots, n; i = 1, \ldots, n\}$ fully determines the price of a Bermudan swaption, the main contribution (up to first order approximation) actually comes from the joint distribution of the co-terminal swap rates at their setting dates $\{y^{i,n+1-i}_{T_i}; i = 1, \ldots, n\}$ (see also Piterbarg [2004]). This is why we are interested in their correlation structure.

3.3 Pricing Bermudan swaptions under the one dimensional swap Markov-functional model

The defining characteristic of the standard Markov-functional model (MF) is that discount bond prices are assumed to be at any time functions of some low-dimensional (usually one or two) Markov process x, which is Markovian in some specified martingale measure. The exact forms are only determined at the exercise dates, i.e.

$D_{T_iT_j}(x_{T_i})$ for $0 \leq i \leq j \leq n$, since this is all that is typically needed in practice. Depending on the application, the functional forms are derived numerically from relevant market prices and the martingale property necessary to maintain the arbitrage-free property of the model. Note that the functional forms of the discount bonds implicitly imply the functional forms of all forward swap/LIBOR rates and vice versa. Given the functional forms, the conditional expected value under the specified martingale measure of a payoff at any exercise date T_i can be derived numerically. Hence, the value of an exotic product can be calculated by backward induction on a grid.

Here, we restrict attention to the development of the one-dimensional swap Markov-functional model (SMF) for the pricing and hedging of Bermudan swaptions. This section starts by reviewing the one-dimensional SMF model and its current choices of driving Markov process. We then propose an alternative choice which is more suitable for our current application.

3.3.1 The one dimensional swap Markov-functional model

In the one-dimensional SMF model, the functional forms of the discount bonds are chosen so that accurate calibration to the market prices of the co-terminal vanilla swaptions is achieved. We assume that these market prices are given by the Black formula with the corresponding co-terminal implied volatilities $\{\widetilde{\sigma}_{i,n+1-i}\}_{i=1,\ldots,n}$. The freedom to specify the law of x allows the model to capture well some features of the real market relevant to the exotic products. For a Bermudan swaption, those features are the correlations of the co-terminal forward swap rates at their setting dates as we discussed in Section 3.2.

In our model, we choose to work with the terminal measure $\mathbb{S}^{n+1}$ which takes the terminal discount bond $D_{\cdot T_{n+1}}$ as the numeraire. Details of the implementation of the SMF model under the terminal measure can be found in Hunt et al. [2000] and Hunt and Kennedy [2004]. We assume the driving process x is a Gaussian process satisfying

$$x_t := \int_0^t \sigma(u) \mathrm{d}W_u,$$

where W denotes a standard Brownian motion under $\mathbb{S}^{n+1}$ and $\sigma(\cdot)$ is a deterministic function of time. For the implementation of the model, we only need to specify the law of x at each exercise date T_i for $i = 1, \ldots, n$.

An important result that was observed in Bennett and Kennedy [2005] is the approximate linear relationship between the logarithms of the co-terminal forward swap rates and x

$$\ln y_t^{i,n+1-i} \approx \underbrace{\gamma_i}_{\text{constant}} x_t + \underbrace{\eta_t^i}_{\text{deterministic}} . \tag{3.3}$$

Consequently, the joint distributions of the log of the co-terminal forward swap rates can be captured via our choice of x since correlation is invariant under the linear transformation

$$\mathrm{Corr}(x_{T_i}, x_{T_j}) = \sqrt{\frac{\xi_{\min(T_i,T_j)}}{\xi_{\max(T_i,T_j)}}} \approx \mathrm{Corr}^{\mathrm{mo}}(\ln y_{T_i}^{i,n+1-i}, \ln y_{T_j}^{j,n+1-j}),$$

where $\xi_{T_i} := \mathrm{Var}(x_{T_i}) = \int_0^{T_i} \sigma^2(t)\mathrm{d}t$ and the superscript "mo" denotes model quantities.

We further note that by matching the model to the Black formula for the co-terminal vanilla swaptions, we have the following approximation in the terminal measure (exact in the associated swaption measure)

$$\mathrm{Var}^{\mathrm{mo}}(\ln y_{T_i}^{i,n+1-i}) \approx \widetilde{\sigma}_{i,n+1-i}^2 T_i. \tag{3.4}$$

Hence, once x is chosen the γ_i's are implicitly determined and from (3.3) and (3.4) we have

$$\gamma_i^2 \xi_{T_i} \approx \widetilde{\sigma}_{i,n+1-i}^2 T_i. \tag{3.5}$$

Note that each γ_i is matched specifically to the associated co-terminal swap rate $y^{i,n+1-i}$ and once the model is calibrated it stays constant from today till expiry. In that sense, the γ_i's are expiry-dependent quantities.

We now present in the following two current candidates for x before exploring other choices in the later sections.

Current candidates:

- **MR:** The first choice is referred to as the mean reversion (MR) driving process with $\sigma(t) = e^{at}$, where $a > 0$ is the mean reversion parameter. It follows that one can write the variance of x at each exercise date T_i as

  $$\xi_{T_i} = \int_0^{T_i} e^{2at}\mathrm{d}t = \frac{1}{2a}(e^{2aT_i} - 1).$$

  For this choice of parametrization, one can see that any changes in the market implied volatilities will not influence x and its auto-correlations once we fix the parameter a. However, the expiry-dependent quantities γ_i's may change as we can see from (3.5). In that sense, the MR process is an example of "parametrization by expiry".

- **HW:** An alternative choice of x is motivated by considering the Hull-White short-rate model which was first introduced in Bennett and Kennedy [2005]. We refer to it as the Hull-White (HW) process. For each $i = 1, \ldots, n$, we have

the following specification for the HW process

$$\xi_{T_i} = \left(\frac{T_{n+1} - T_i}{(1 + \alpha_i y_0^{i,n+1-i})(\psi_{T_{n+1}} - \psi_{T_i})} \right)^2 \widetilde{\sigma}^2_{i,n+1-i} T_i, \tag{3.6}$$

where $\psi_{T_i} = \frac{1}{a}(1 - e^{-aT_i}), a > 0$. In contrast to the MR process, any changes in the market co-terminal implied volatilities will have an immediate effect on x and its auto-correlations. From the linear approximation in (3.3), we see that the instantaneous volatilities of the co-terminal swap rates will be altered in certain time periods. On the other hand, the expiry-dependent quantities γ_i's will stay the same as we can see from substitution of the expression (3.6) into (3.5). In that sense, the HW process is an example of "parametrization by time".

For both the above model parametrizations, an increase in the implied volatility of one of the co-terminal swap rates tends to increase the value of a Bermudan. This is not surprising as the value of the associated vanilla swaption has increased. But the optionality of a Bermudan provides extra value in addition to the value of the underlying vanilla options. This extra value is highly dependent on the correlations between the co-terminal swap rates at their setting dates and for the above two models these correlations behave very differently in response to changes in the co-terminal implied volatilities. This leads to very different hedging profiles as we shall see in the later sections. In the next section we investigate the essential difference in nature between the two parametrizations by considering the underlying LIBORs.

3.3.2 Parametrizations by time and by expiry

In the previous subsection, we discussed the idea of parametrizations by expiry and by time in terms of the responses of the instantaneous volatilities of the co-terminal swap rates to a shift in the implied volatilities. Here we explore how this idea carries over to the LIBORs as they are the basic building blocks of any interest rate model.

For all choices of x, the linear approximation in (3.3) implies that the instantaneous volatility of the log of the co-terminal forward swap rate $y_t^{i,n+1-i}$ is approximately $\gamma_i \sigma(t)$ under the terminal measure. In order to gain insight into the effect of shifting the implied volatilities, we make the simplifying assumption that each log-LIBOR $\ln L^i$ has a positive and deterministic volatility function $\sigma_i(t), t \leq T_i$. Under this assumption, we can use the approximation described in Appendix 3.A. In a one factor model instead of the multi-factor setting in Appendix 3.A, the instantaneous volatility of the log of the co-terminal forward swap rate $y^{i,n+1-i}$ can be linked to

the instantaneous volatilities of the log-LIBORs by the following approximation

$$\gamma_i \sigma(t) \approx \sum_{k=i}^{n} \zeta_k^{i,n+1-i}(0)\sigma_k(t), \tag{3.7}$$

where $\{\zeta_k^{i,n+1-i}(0)\}_{k=i,\ldots,n}$ are constant empirical weights that depend on the initial discount curve. This can be seen from SDE (3.36) in Appendix 3.A.2.

Since the last LIBOR $L^n = y^{n,1}$, we have that $\sigma_n(t) \approx \gamma_n \sigma(t)$. Using the derived form for $\sigma_n(t)$, we can deduce $\sigma_{n-1}(t)$ by considering the approximation in (3.7) for $\gamma_{n-1}\sigma(t)$

$$\sigma_{n-1}(t) \approx \sigma_{n-1}\sigma(t),$$

where we let

$$\sigma_{n-1} := \frac{\gamma_{n-1} - \zeta_n^{n-1,2}(0)\gamma_n}{\zeta_{n-1}^{n-1,2}(0)}.$$

We assume $\sigma_{n-1} > 0$ so that $\sigma_{n-1}(\cdot)$ will also be positive as we assumed earlier. Inductively, assume we have that $\sigma_k(t) \approx \sigma_k \sigma(t)$ where σ_k is a positive constant for each $k = i+1, \ldots, n$. When $k = n$, σ_n is the same as γ_n. We can then derive $\sigma_i(t)$ by considering the approximation in (3.7) for $\gamma_i\sigma(t)$. We rewrite (3.7) in the following form

$$\begin{aligned} \gamma_i \sigma(t) &\approx \zeta_i^{i,n+1-i}(0)\sigma_i(t) + \sum_{k=i+1}^{n} \zeta_k^{i,n+1-i}(0)\sigma_k\sigma(t) \\ \iff \sigma_i(t) &\approx \left(\frac{\gamma_i - \sum_{k=i+1}^{n} \zeta_k^{i,n+1-i}(0)\sigma_k}{\zeta_i^{i,n+1-i}(0)} \right) \sigma(t). \end{aligned}$$

This again reduces $\sigma_i(t)$ approximately to the form $\sigma_i\sigma(t)$ where

$$\sigma_i := \frac{\gamma_i - \sum_{k=i+1}^{n} \zeta_k^{i,n+1-i}(0)\sigma_k}{\zeta_i^{i,n+1-i}(0)}. \tag{3.8}$$

This concludes that $\sigma_i(t) \approx \sigma_i\sigma(t)$ for all $i = 1, \ldots, n$ where each constant σ_i can be derived inductively by (3.8) and is assumed to be positive. One can see that each σ_i depends on $\{\gamma_k\}_{k=i,\ldots,n}$ and the empirical weights $\{\zeta_k^{i,n+1-i}(0)\}_{k=i,\ldots,n}$, $\{\zeta_k^{i+1,n-i}(0)\}_{k=i+1,\ldots,n}$, $\ldots$, $\{\zeta_k^{n-1,2}(0)\}_{k=n-1,n}$. Since these empirical weights do not depend on the implied volatilities, we will safely ignore their involvements in the next discussion.

We now analyze how shifting the implied volatilities affects the instantaneous volatility functions $\{\sigma_i(\cdot)\}_{i=1,\ldots,n}$ of the log-LIBORs for each choice of x.

Parametrization by expiry: For the MR process, by the approximation in (3.5)

we have that

$$\gamma_i \approx \frac{\widetilde{\sigma}_{i,n+1-i}\sqrt{T_i}}{\sqrt{\frac{e^{2aT_i}-1}{2a}}}, \qquad i = 1, \ldots, n. \tag{3.9}$$

Suppose we want to bump the co-terminal implied volatility $\widetilde{\sigma}_{i,n+1-i}$ and keep the rest the same. It is clear that the instantaneous volatility $\sigma(t) = e^{at}$ of the MR process will not be affected. We observe other effects and summarize them below:

- γ_j for $j \neq i$ are unchanged as we can see from (3.9). This then follows from (3.8) that the constants $\{\sigma_j\}_{j=i+1,\ldots,n}$ and hence $\{\sigma_j(\cdot)\}_{j=i+1,\ldots,n}$ also remain unchanged.
- γ_i will increase directly as a result of (3.9). From (3.8), we see that σ_i and hence $\sigma_i(\cdot)$ will increase as $\{\sigma_j\}_{j=i+1,\ldots,n}$ are unchanged.
- Since γ_{i-1} and $\{\sigma_j\}_{j=i+1,\ldots,n}$ stay the same but σ_i increases, again we can see from (3.8) that σ_{i-1} and hence $\sigma_{i-1}(\cdot)$ will decrease.
- The effects on $\{\sigma_k\}_{k=1,\ldots,i-2}$ and hence $\{\sigma_k(\cdot)\}_{k=1,\ldots,i-2}$ will be quite small. This is because the increase in σ_i and decrease in σ_{i-1} tend to cancel each other out in the sum in (3.8) when we consider σ_k for $k < i-1$.

Note that all the above effects are (global) from today till expiry (illustrated in figure 3.1). In that sense, the instantaneous volatilities of the log-LIBORs are clearly parameterized by expiry.

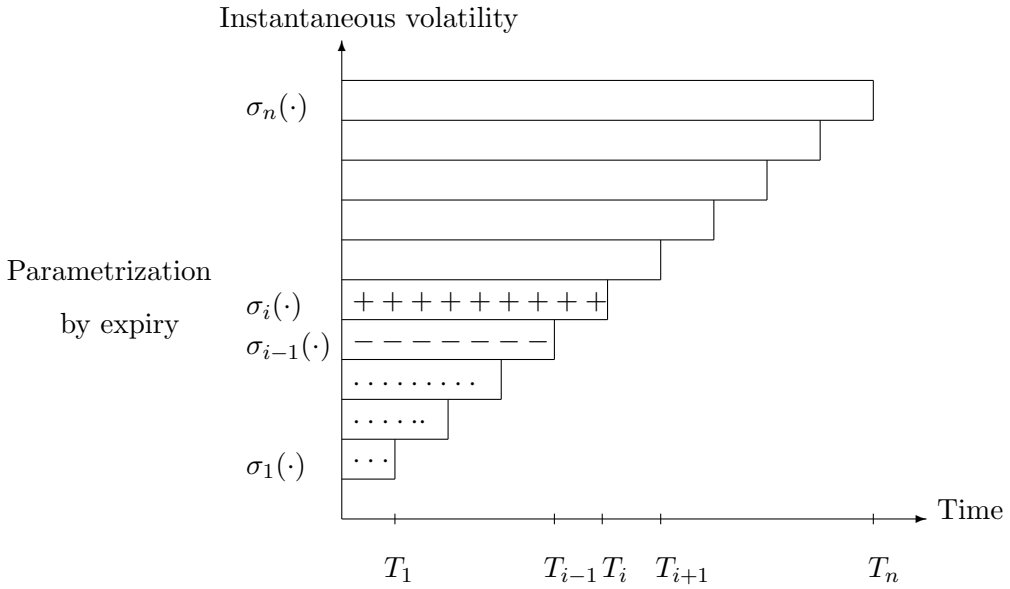


Figure 3.1: Global effect of bumping $\widetilde{\sigma}_{i,n+1-i}$ on the instantaneous volatility functions of the log-LIBORs. The dots represent a very small effect.

Parametrization by time: For the HW process where ξ_{T_i} is given by (3.6), it can be seen from (3.5) that the γ_i's are independent of the implied volatilities

$$\gamma_i \approx \frac{(1+\alpha_i y_0^{i,n+1-i})(\psi_{T_{n+1}} - \psi_{T_i})}{T_{n+1} - T_i}, \qquad \psi_{T_i} = \frac{1}{a}(1 - e^{-aT_i}), \qquad a > 0. \tag{3.10}$$

Although the instantaneous volatility function $\sigma(\cdot)$ of x is not defined explicitly, we know it exists such that $\int_0^{T_i} \sigma^2(t)\mathrm{d}t = \xi_{T_i}$ is given by (3.6). This is because this form of ξ_{T_i} is strictly increasing in T_i.

We now bump the co-terminal implied volatility $\widetilde{\sigma}_{i,n+1-i}$ and keep the rest unchanged. It is clear from (3.10) that γ_j will stay the same for $j = 1, \ldots, n$. Hence, it follows from (3.8) that the constants σ_j will also remain unchanged for all j. The only effect of the bump is on the function $\sigma(\cdot)$ (see (3.6)). One can see that the variance of x at T_i is shifted but those at the other exercise dates remain unchanged. Consequently, we have that the only effect on x is the following

$$\begin{aligned} \xi_{T_i} - \xi_{T_{i-1}} &= \int_{T_{i-1}}^{T_i} \sigma^2(t)\mathrm{d}t \qquad \text{increases,} \\ \xi_{T_{i+1}} - \xi_{T_i} &= \int_{T_i}^{T_{i+1}} \sigma^2(t)\mathrm{d}t \qquad \text{decreases.} \end{aligned}$$

The above effect implies that on average the instantaneous volatility function $\sigma(\cdot)$ of x is increased during the time period $(T_{i-1}, T_i]$ but is decreased during the next one $(T_i, T_{i+1}]$. Since $\sigma_j(t) \approx \sigma_j \sigma(t)$ is only defined during the corresponding LIBOR's life, i.e. $t \in [0, T_j]$, the effect on $\sigma(\cdot)$ only carries over to $\{\sigma_j(\cdot)\}_{j=i,\ldots,n}$. It is then clear that on average the collection of instantaneous volatilities $\{\sigma_j(\cdot)\}_{j=i,\ldots,n}$ will increase and decrease during the two consecutive time intervals $(T_{i-1}, T_i]$ and $(T_i, T_{i+1}]$ respectively (figure 3.2). For the last co-terminal implied volatility $\widetilde{\sigma}_{n,1}$, the equivalent effect is that $\sigma_n(\cdot)$ will only increase during the period $(T_{n-1}, T_n]$. Note that the above effects are local as the instantaneous volatilities of the log-LIBORs are only shocked locally for some particular time periods in response to the movement of the corresponding implied volatility. In that sense, the instantaneous volatilities of the log-LIBORs are clearly parameterized by time.

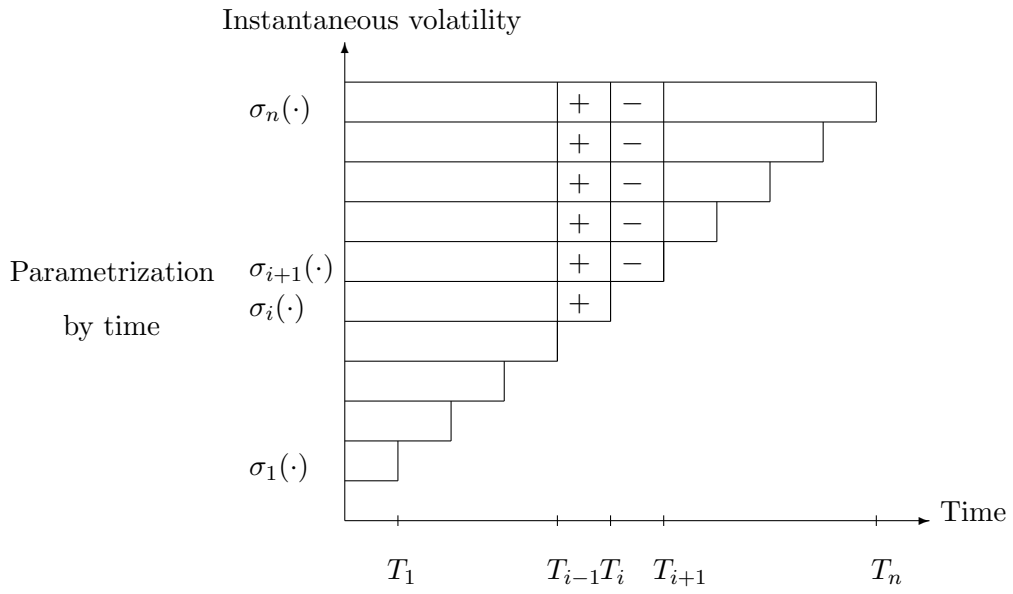


Figure 3.2: Local effects of bumping the $\widetilde{\sigma}_{i,n+1-i}$ on the instantaneous volatility functions of the log-LIBORs.

The difference in parametrizations mentioned above has a fundamental effect on the hedging of a Bermudan swaption. Specifically, the global and local effects of the parametrizations by expiry and by time influence the correlations of the forward LIBORs and the co-terminal forward swap rates in very different ways. This fact, in turn, leads to very different hedging behaviours of correlation-dependent products like the Bermudan swaption. We will investigate further the difference in their vega profiles in Section 3.4.2. Since the "parametrization by time" outperforms the other type as we explore later in Section 3.5, we will next propose an alternative for this parametrization.

3.3.3 An alternative parametrization of time

We recall that the correlation of the MR process is fixed from the outset while the HW specification links the correlation structure of the model/co-terminal swap rates explicitly to the market implied volatilities. However at each exercise date T_i, the HW process only takes into account the co-terminal implied volatility $\widetilde{\sigma}_{i,n+1-i}$. In this section, we explore alternative ways to specify the x process which link the model's correlation structure to implied volatilities of different tenors (see table 3.1).

Tenor		1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
Expiry	1	…	…	…	…	…	…	…	…	…	$\widetilde{\sigma}_{1,10}$	$\widetilde{\sigma}_{1,1}$	$\widetilde{\sigma}_{1,2}$	$\widetilde{\sigma}_{1,3}$	$\widetilde{\sigma}_{1,4}$	$\widetilde{\sigma}_{1,5}$	$\widetilde{\sigma}_{1,6}$	$\widetilde{\sigma}_{1,7}$	$\widetilde{\sigma}_{1,8}$	$\widetilde{\sigma}_{1,9}$	$\widetilde{\sigma}_{1,10}$
	2	…	…	…	…	…	…	…	…	$\widetilde{\sigma}_{2,9}$	…	$\widetilde{\sigma}_{2,1}$	$\widetilde{\sigma}_{2,2}$	$\widetilde{\sigma}_{2,3}$	$\widetilde{\sigma}_{2,4}$	$\widetilde{\sigma}_{2,5}$	$\widetilde{\sigma}_{2,6}$	$\widetilde{\sigma}_{2,7}$	$\widetilde{\sigma}_{2,8}$	$\widetilde{\sigma}_{2,9}$	…
	3	…	…	…	…	…	…	…	$\widetilde{\sigma}_{3,8}$	…	…	$\widetilde{\sigma}_{3,1}$	$\widetilde{\sigma}_{3,2}$	$\widetilde{\sigma}_{3,3}$	$\widetilde{\sigma}_{3,4}$	$\widetilde{\sigma}_{3,5}$	$\widetilde{\sigma}_{3,6}$	$\widetilde{\sigma}_{3,7}$	$\widetilde{\sigma}_{3,8}$	…	…
	4	…	…	…	…	…	…	$\widetilde{\sigma}_{4,7}$	…	…	…	$\widetilde{\sigma}_{4,1}$	$\widetilde{\sigma}_{4,2}$	$\widetilde{\sigma}_{4,3}$	$\widetilde{\sigma}_{4,4}$	$\widetilde{\sigma}_{4,5}$	$\widetilde{\sigma}_{4,6}$	$\widetilde{\sigma}_{4,7}$	…	…	…
	5	…	…	…	…	…	$\widetilde{\sigma}_{5,6}$	…	…	…	…	$\widetilde{\sigma}_{5,1}$	$\widetilde{\sigma}_{5,2}$	$\widetilde{\sigma}_{5,3}$	$\widetilde{\sigma}_{5,4}$	$\widetilde{\sigma}_{5,5}$	$\widetilde{\sigma}_{5,6}$	…	…	…	…
	6	…	…	…	…	$\widetilde{\sigma}_{6,5}$	…	…	…	…	…	$\widetilde{\sigma}_{6,1}$	$\widetilde{\sigma}_{6,2}$	$\widetilde{\sigma}_{6,3}$	$\widetilde{\sigma}_{6,4}$	$\widetilde{\sigma}_{6,5}$	…	…	…	…	…
	7	…	…	…	$\widetilde{\sigma}_{7,4}$	…	…	…	…	…	…	$\widetilde{\sigma}_{7,1}$	$\widetilde{\sigma}_{7,2}$	$\widetilde{\sigma}_{7,3}$	$\widetilde{\sigma}_{7,4}$	…	…	…	…	…	…
	8	…	…	$\widetilde{\sigma}_{8,3}$	…	…	…	…	…	…	…	$\widetilde{\sigma}_{8,1}$	$\widetilde{\sigma}_{8,2}$	$\widetilde{\sigma}_{8,3}$	…	…	…	…	…	…	…
	9	…	$\widetilde{\sigma}_{9,2}$	…	…	…	…	…	…	…	…	$\widetilde{\sigma}_{9,1}$	$\widetilde{\sigma}_{9,2}$	…	…	…	…	…	…	…	…
	10	$\widetilde{\sigma}_{10,1}$	…	…	…	…	…	…	…	…	…	$\widetilde{\sigma}_{10,1}$	…	…	…	…	…	…	…	…	…

Table 3.1: Market data from the swaption matrix to be incorporated into the driving process x for a 11 years annual Bermudan swaption. HW's approach (left), alternative approach (right).

3.3.3.1 One step covariance

One way to view a Bermudan swaption is as the right to choose between the associated European swaptions. In setting up a model, one might choose to try to capture the correlations between the co-terminal swap rates at their setting dates. These are the correlations that matter when pricing a Bermudan swaption. In a one factor model, we cannot capture all these correlations. One choice is to consider the one step correlations, i.e. the correlation of $y_{T_i}^{i,n+1-i}$ with its nearest neighbour $y_{T_{i+1}}^{i+1,n-i}$ for each i. Note that we will work with the log of the swap rates because it allows for a direct connection with the driving process x as we shall see in the model's setup.

In the approach adopted here, we estimate the one step covariances, i.e. $\mathrm{Cov}(\ln y_{T_i}^{i,n+1-i}, \ln y_{T_{i+1}}^{i+1,n-i})$ for $i = 1, \ldots, n-1$, using the swaption matrix from the market. This is a two-step procedure. The first step is to approximate the correlations of the log-LIBORs at each exercise date by a global fit to the swaption matrix. The second step is to deduce the corresponding covariances of the log-LIBORs by performing a local fit to each row of the swaption matrix and using the correlations from the first step. We then use these covariances to derive the required one step covariances (see equation (3.37)). In fact, we only need $\mathrm{Cov}(\ln y_{T_i}^{i,n+1-i}, \ln y_{T_i}^{i+1,n-i})$ as

$$\mathrm{Cov}(\ln y_{T_i}^{i,n+1-i}, \ln y_{T_j}^{j,n+1-j}) \approx \mathrm{Cov}(\ln y_{T_i}^{i,n+1-i}, \ln y_{T_i}^{j,n+1-j}),$$

for $i < j \leq n$ (see Appendix 3.A.2). Details of the whole approximation procedure can be found in Appendix 3.A. In what follows we will use the superscript "ma" to denote quantities estimated from the market.

Model's setup: Once we have estimated the one step covariances from the market, we set up the model as follows. Recall the linear approximation under the

terminal measure in Bennett and Kennedy [2005], for $i < n$

$$\begin{aligned} \gamma_i x_t + \eta_t^i &\approx \ln y_t^{i,n+1-i} \\ \Rightarrow \mathrm{Corr}(x_{T_i}, x_{T_{i+1}}) &\approx \mathrm{Corr}^{\mathrm{mo}}(\ln y_{T_i}^{i,n+1-i}, \ln y_{T_{i+1}}^{i+1,n-i}) \\ \iff \sqrt{\frac{\xi_{T_i}}{\xi_{T_{i+1}}}} &\approx \frac{\mathrm{Cov}^{\mathrm{mo}}(\ln y_{T_i}^{i,n+1-i}, \ln y_{T_i}^{i+1,n-i})}{\sqrt{\mathrm{Var}^{\mathrm{mo}}(\ln y_{T_i}^{i,n+1-i})\mathrm{Var}^{\mathrm{mo}}(\ln y_{T_{i+1}}^{i+1,n-i})}}. \end{aligned} \tag{3.11}$$

As for each $i = 1, \ldots, n$ $\mathrm{Var}^{\mathrm{mo}}(\ln y_{T_i}^{i,n+1-i})$ can be inferred from the corresponding Black implied volatility $\widetilde{\sigma}_{i,n+1-i}$, we can now incorporate the one step covariances into the model as described below.

- Without loss of generality, fix $\xi_{T_n} = \widetilde{\sigma}_{n,1}^2 T_n$.
- At T_{n-1}, by knowledge of $\mathrm{Cov}^{\mathrm{ma}}(\ln y_{T_{n-1}}^{n-1,2}, \ln y_{T_n}^{n,1})$ and hence the one step correlation $\mathrm{Corr}^{\mathrm{ma}}(\ln y_{T_{n-1}}^{n-1,2}, \ln y_{T_n}^{n,1}) = \frac{\mathrm{Cov}^{\mathrm{ma}}(\ln y_{T_{n-1}}^{n-1,2}, \ln y_{T_n}^{n,1})}{\widetilde{\sigma}_{n-1,2}\sqrt{T_{n-1}}\widetilde{\sigma}_{n,1}\sqrt{T_n}}$ that we have estimated from the market, we can recover $\xi_{T_{n-1}}$ by fixing

$$\begin{aligned} \sqrt{\frac{\xi_{T_{n-1}}}{\xi_{T_n}}} &= \mathrm{Corr}^{\mathrm{ma}}(\ln y_{T_{n-1}}^{n-1,2}, \ln y_{T_n}^{n,1}) \\ \iff \sqrt{\xi_{T_{n-1}}} &= \mathrm{Corr}^{\mathrm{ma}}(\ln y_{T_{n-1}}^{n-1,2}, \ln y_{T_n}^{n,1})\sqrt{\xi_{T_n}}, \end{aligned}$$

  where we use the relation in (3.11).

- Inductively, assume we are at T_i and have derived ξ_{T_j} for $j > i$ from the previous steps. By the approximation for $\mathrm{Corr}^{\mathrm{ma}}(\ln y_{T_i}^{i,n+1-i}, \ln y_{T_{i+1}}^{i+1,n-i})$ from the market and the knowledge of $\xi_{T_{i+1}}$, we can fix ξ_{T_i}

$$\sqrt{\xi_{T_i}} = \mathrm{Corr}^{\mathrm{ma}}(\ln y_{T_i}^{i,n+1-i}, \ln y_{T_{i+1}}^{i+1,n-i})\sqrt{\xi_{T_{i+1}}},$$

  where we again use (3.11).

- We have now fixed ξ_{T_i} for $i = 1, \ldots, n$ and the SMF model can be implemented on the grid.

The above model is an example of parametrization by time and its overall vega profile (in terms of a Bermudan swaption) has a close connection to that of the HW model as we shall see in Section 3.4. As implied volatilities change, the implied correlations in the market change. The one step covariance model attempts to hedge this risk but with the focus just on the one step covariance with the next co-terminal swap rate. Clearly, we are ignoring some important market information. This is exactly the motivation for our next proposed model.

3.3.3.2 Weighted covariance

We propose another alternative choice of x based on the following intuition. In order to make a decision whether to exercise the option at T_i, the one step correlation $\text{Corr}(\ln y_{T_i}^{i,n+1-i}, \ln y_{T_{i+1}}^{i+1,n-i})$ will be the most important as T_{i+1} is immediately after T_i. The correlations with the later rates are less significant since we can wait until later to decide.

For each $i = 1, \ldots, n-1$, in addition to the one step covariance we can choose to take into account the significance of the covariances $\text{Cov}(\ln y_{T_i}^{i,n+1-i}, \ln y_{T_j}^{j,n+1-j})$ for all $j > i$ but with different levels of impact on the Bermudan swaption's price. Again, as there is only one factor in the model we cannot capture everything. One way is to consider the weighted covariance $\text{Cov}(\ln y_{T_i}^{i,n+1-i}, \sum_{j=i+1}^{n} p^{T_j-T_i} \ln y_{T_j}^{j,n+1-j})$ at each exercise date T_i where the weight is chosen to be a monotonically decreasing function in $T_j - T_i$

$$p^{T_j-T_i} = \exp[-\alpha(T_j - T_i)], \qquad j > i,$$

for some $\alpha > 0$. Note that the weighted covariance can be estimated in exactly the same way as we estimate the one step covariance from the market (see Appendix 3.A). As we shall see in Section 3.4.3, the weighted covariance model spreads the vega responses over the swaption matrix while the one step covariance model only assigns a significant contribution to the first column and the reverse diagonal. It is this feature that gives the weighted covariance model a potential hedging advantage over the one step covariance model.

Model's setup: For $i = 1, \ldots, n-1$, we calibrate the model to the following market quantity

$$\text{Cov}^{\text{ma}}(\ln y_{T_i}^{i,n+1-i}, \sum_{j=i+1}^{n} p^{T_j-T_i} \ln y_{T_j}^{j,n+1-j}) \approx \sum_{j=i+1}^{n} p^{T_j-T_i} \text{Cov}^{\text{ma}}(\ln y_{T_i}^{i,n+1-i}, \ln y_{T_i}^{j,n+1-j}). \tag{3.12}$$

For ease of exposition, we denote this market quantity by B_i. One can incorporate the B_i's into the model as follows

- Without loss of generality, fix $\xi_{T_n}^{\alpha} = \widetilde{\sigma}_{n,1}^2 T_n$[1].
- At T_{n-1}, we only need to estimate from the market the one step covariance $\text{Cov}^{\text{ma}}(\ln y_{T_{n-1}}^{n-1,2}, \ln y_{T_{n-1}}^{n,1})$ for B_{n-1} and consequently the one step correlation $\text{Corr}^{\text{ma}}(\ln y_{T_{n-1}}^{n-1,2}, \ln y_{T_n}^{n,1})$. Since we want to calibrate the model's correlation

[1]The superscript α is added to emphasize the dependence.

structure to B_{n-1}, we then require that

$$\begin{aligned} p\sqrt{\frac{\xi^{\alpha}_{T_{n-1}}}{\xi^{\alpha}_{T_n}}}\widetilde{\sigma}_{n-1,2}\sqrt{T_{n-1}}\widetilde{\sigma}_{n,1}\sqrt{T_n} &= B_{n-1} \\ \iff \sqrt{\xi^{\alpha}_{T_{n-1}}} &= \mathrm{Corr}^{\mathrm{ma}}(\ln y^{n-1,2}_{T_{n-1}}, \ln y^{n,1}_{T_n})\sqrt{\xi^{\alpha}_{T_n}}, \end{aligned}$$

where we use the relation in (3.11). Note that this step is the same as that in the one step covariance model.

- Inductively, assume we are at T_i and have derived $\{\xi^{\alpha}_{T_j}\}_{j=i+1,\ldots,n}$ from the previous steps, $\xi^{\alpha}_{T_i}$ will then be obtained by backward induction. Since we have estimated B_i from the market and want to calibrate the model's correlation structure to B_i, we require that

$$\sum_{j=i+1}^{n} p^{T_j-T_i}\sqrt{\frac{\xi^{\alpha}_{T_i}}{\xi^{\alpha}_{T_j}}}\widetilde{\sigma}_{i,n+1-i}\sqrt{T_i}\widetilde{\sigma}_{j,n+1-j}\sqrt{T_j} = B_i, \tag{3.13}$$

  where we use $\mathrm{Corr}^{\mathrm{ma}}(\ln y^{i,n+1-i}_{T_i}, \ln y^{j,n+1-j}_{T_j})$ for $j > i$ instead of just $j = i+1$ as in the one step covariance case. Hence, from (3.13) we fix

$$\sqrt{\xi^{\alpha}_{T_i}} = \frac{B_i}{\widetilde{\sigma}_{i,n+1-i}\sqrt{T_i}\sum_{j=i+1}^{n} p^{T_j-T_i}\sqrt{\frac{1}{\xi^{\alpha}_{T_j}}}\widetilde{\sigma}_{j,n+1-j}\sqrt{T_j}}. \tag{3.14}$$

- We have now fixed $\xi^{\alpha}_{T_i}$ for $i = 1, \ldots, n$ and the SMF model can be implemented on the grid.

One can immediately see that the one step covariance process is a special case of the general weighted covariance process when α is very large. The reason is the following. When α is sufficiently large, $p^{T_j-T_i}$ will decay exponentially fast and all the weights will then become insignificant compared with $p^{T_{i+1}-T_i}$. Consequently, only the one step covariance matters in the market quantity B_i and the weighted covariance process is reduced to the one step covariance one.

Remark 1 *For both the one step and weighted covariance models, one can view the vectors of Black implied volatilities* $(\widetilde{\sigma}_{i,1}, \widetilde{\sigma}_{i,2}, \ldots, \widetilde{\sigma}_{i,n+1-i})$ *for* $i = 1, \ldots, n$ *as the model's initial inputs (see the global and local fits in Appendix 3.A). It follows from the constructions of both the one step and weighted covariance processes that one can write*

$$\xi^{\alpha}_{T_i} = f^i(\xi^{\alpha}_{T_{i+1}}, \ldots, \xi^{\alpha}_{T_n}; \{\widetilde{\sigma}_{i,k}\}_{k=1,\ldots,n+1-i}; \{\widetilde{\sigma}_{j,n+1-j}\}_{j=i+1,\ldots,n}), \tag{3.15}$$

where f^i *is some deterministic function (see Appendix 3.A.2 for more details). This*

cascade structure of the model clearly has an important implication on the response of the Bermudan price to changes in implied volatilities.

Although the one step and weighted covariance processes appear to be more complicated than the HW process, it turns out that they are still quite similar in some context. In the next section, we will explore further their connection through the Bermudan swaption's vegas.

3.4 Vegas

In this section, we study the vegas of a Bermudan swaption produced by the different models. While the deltas and the gammas do not vary so much from model to model as we shall see in Section 3.5, the vegas prove to be the most influential in the hedging of a Bermudan. In addition, the underlying parametrization of the model has an important implication for the vegas. It is, therefore, worthwhile to investigate the behaviour of the vegas from different perspectives to explore the model's structure. In Subsection 3.4.1, we review analytically the vegas under different models. We then study the vegas numerically and investigate further the link between them.

3.4.1 The vega computation under the swap Markov-functional model

In order to compute the price of a Bermudan swaption in a SMF model, we need two sets of initial inputs and a covariance/correlation structure (captured through the process x). The first set of inputs are the initial discount bonds which can be safely ignored in this discussion as we are only concerned with the vegas here. The second set of inputs are the co-terminal implied volatilities which are used to recover the prices of the underlying co-terminal vanilla swaptions and fix the functional forms of the corresponding co-terminal forward swap rates at their setting dates. In the implementation of the SMF model, this is the calibration to the market marginals and it is done for all different specifications of x. In general, one can view the value of a Bermudan $\hat{V}_{T_0}$ as a price function which maps (the square root of) the variances of x and the second set of inputs to a real positive value:

$$\begin{aligned} \hat{V}_{T_0} : \mathbb{R}^n \times \mathbb{R}^n &\rightarrow \mathbb{R}^+ \\ \hat{V}_{T_0}(\overline{\xi}, \overline{\sigma}) &:= v_0, \end{aligned} \tag{3.16}$$

where $\overline{\xi} = (\sqrt{\xi_{T_1}}, \sqrt{\xi_{T_2}}, \ldots, \sqrt{\xi_{T_n}})$ and $\overline{\sigma} = (\widetilde{\sigma}_{1,n}, \widetilde{\sigma}_{2,n-1}, \ldots, \widetilde{\sigma}_{n,1})$. For the weighted covariance process, the equivalent input from x is $\overline{\xi}^{\alpha} = (\sqrt{\xi^{\alpha}_{T_1}}, \sqrt{\xi^{\alpha}_{T_2}}, \ldots, \sqrt{\xi^{\alpha}_{T_n}})$. Note that the notation $\overline{\xi}^{\alpha}$ when $\alpha \rightarrow \infty$ indicates the one step covariance case as we explained in Section 3.3.3.2. For the data we are working with, it is observed that

the vectors $\overline{\xi}^{\alpha}$ are quite similar for all α and hence different choices for α result in similar prices for the Bermudan.

Define the vega $\nu_{i,k}$ to be the total derivative of the Bermudan swaption's price with respect to $\widetilde{\sigma}_{i,k}$ for each $i = 1, \ldots, n$ and $k = 1, \ldots, n+1-i$:

$$\nu_{i,k} := \frac{\mathrm{d}\hat{V}_{T_0}}{\mathrm{d}\widetilde{\sigma}_{i,k}}. \tag{3.17}$$

We apply the finite difference/bumping-revaluation method to calculate these derivatives numerically.

We now consider the vegas on a particular i$^{\text{th}}$ row of the swaption matrix. In order to distinguish the vegas produced by different models, we denote $\nu^{\alpha}_{i,k}$ for the weighted covariance process. Note again that the notation $\nu^{\alpha}_{i,k}$ as $\alpha \to \infty$ indicates the vegas for the one step covariance process. For the HW and MR models, we denote the vegas by $\nu^{\text{hw}}_{i,k}$ and $\nu^{\text{mr}}_{i,k}$ respectively. Note that only the co-terminal vegas $\nu^{\text{hw}}_{i,n+1-i}$ and $\nu^{\text{mr}}_{i,n+1-i}$ matter in the HW and MR models which follows directly from their setups. The other vegas are all zero for these models.

1. $k = 1, \ldots, n-i$ (**off reverse diagonal**): For the one step and weighted covariance models, as $\widetilde{\sigma}_{i,k}$ is only involved in $\overline{\xi}^{\alpha}$, by the chain rule and equation (3.15) we have that

$$\nu^{\alpha}_{i,k} = \frac{\mathrm{d}\hat{V}_{T_0}}{\mathrm{d}\widetilde{\sigma}_{i,k}} = \sum_{s=1}^{i} \frac{\partial \hat{V}_{T_0}}{\partial \sqrt{\xi^{\alpha}_{T_s}}} \times \frac{\mathrm{d}\sqrt{\xi^{\alpha}_{T_s}}}{\mathrm{d}\widetilde{\sigma}_{i,k}}. \tag{3.18}$$

   The term $\frac{\partial \hat{V}_{T_0}}{\partial \sqrt{\xi^{\alpha}_{T_s}}}$ should be interpreted as the partial derivative of the price function $\hat{V}_{T_0}$ with respect to the s$^{\text{th}}$ coordinate of the vector $\overline{\xi}^{\alpha}$. For the total derivatives $\frac{\mathrm{d}\sqrt{\xi^{\alpha}_{T_s}}}{\mathrm{d}\widetilde{\sigma}_{i,k}}$ where $1 \le s < i$, by equation (3.15) it is clear that the dependence of $\sqrt{\xi^{\alpha}_{T_s}}$ on $\widetilde{\sigma}_{i,k}$ is through $\{\sqrt{\xi^{\alpha}_{T_{j^*}}}\}_{j^*=s+1,\ldots,i}$.

   Note that the price of a Bermudan swaption is not sensitive to these implied volatility inputs under the HW and MR models. Hence, as noted above we have zero values for the corresponding vegas in these models.

2. $k = n+1-i$ (**reverse diagonal**): For the one step and weighted covariance models, as $\widetilde{\sigma}_{i,n+1-i}$ is involved in both $\overline{\xi}^{\alpha}$ and $\overline{\sigma}$ we have that

$$\nu^{\alpha}_{i,n+1-i} = \frac{\partial \hat{V}_{T_0}}{\partial \widetilde{\sigma}_{i,n+1-i}} + \sum_{s=1}^{i} \frac{\partial \hat{V}_{T_0}}{\partial \sqrt{\xi^{\alpha}_{T_s}}} \times \frac{\mathrm{d}\sqrt{\xi^{\alpha}_{T_s}}}{\mathrm{d}\widetilde{\sigma}_{i,n+1-i}}. \tag{3.19}$$

   Note that the first term on the right hand side of (3.19) is approximately the i$^{\text{th}}$ bucket vega $\nu^{\text{mr}}_{i,n+1-i}$ of the MR process when the prices are comparable

between models. This term only reflects the change of the marginal distribution of $y_{T_i}^{i,n+1-i}$ under its own swaption measure. For the HW process, the equivalent vega is

$$\begin{aligned} \nu_{i,n+1-i}^{\text{hw}} &= \frac{\partial \hat{V}_{T_0}}{\partial \widetilde{\sigma}_{i,n+1-i}} + \frac{\partial \hat{V}_{T_0}}{\partial \sqrt{\xi_{T_i}}} \times \frac{\mathrm{d}\sqrt{\xi_{T_i}}}{\mathrm{d}\widetilde{\sigma}_{i,n+1-i}} \\ &\approx \nu_{i,n+1-i}^{\text{mr}} + \frac{\partial \hat{V}_{T_0}}{\partial \sqrt{\xi_{T_i}}} \times \frac{\mathrm{d}\sqrt{\xi_{T_i}}}{\mathrm{d}\widetilde{\sigma}_{i,n+1-i}}. \end{aligned} \tag{3.20}$$

One can see that the above differences between $\nu_{i,n+1-i}^{\text{hw}}$ and $\nu_{i,n+1-i}^{\alpha}$ on one hand and $\nu_{i,n+1-i}^{\text{hw}}$ and $\nu_{i,n+1-i}^{\text{mr}}$ on the other hand come from the differences in their correlation structures. Since the HW and the one step and the weighted covariance models are different examples of parametrization by time, their vegas are closely connected as we shall see in Section 3.4.4. For the MR and HW models which respond to the reverse diagonal only, the difference in their correlation structures follow directly from the difference between the parameterizations by expiry and by time. We will review this difference in the next section.

3.4.2 The Bermudan swaption's vegas under the HW and MR models

The example we consider here is a 11 years annual payer Bermudan swaption with fixed rate $K = 5\%$, notional $N = 100$ million and the following initial data:

Tenor		1	2	3	4	5	6	7	8	9	10
Expiry	1	13.12	13.19	13.21	13.21	13.22	13.00	12.78	12.58	12.39	12.17
	2	13.16	13.16	13.09	13.04	12.92	12.72	12.51	12.31	12.12	...
	3	13.06	12.97	12.91	12.82	12.66	12.42	12.20	12.03	...	...
	4	12.95	12.82	12.72	12.51	12.32	12.14	11.93	...	...	...
	5	12.76	12.57	12.43	12.24	12.03	11.87	...	...	...	...
	6	12.38	12.19	12.12	11.89	11.71	...	...	...	...	...
	7	12.10	11.89	11.77	11.59	...	...	...	...	...	...
	8	11.69	11.56	11.46	...	...	...	...	...	...	...
	9	11.48	11.31	...	...	...	...	...	...	...	...
	10	11.19	...	...	...	...	...	...	...	...	...

Table 3.2: Black implied volatilities (%) of the ATM swaptions on October 17, 2007.

and

Tenor		1	2	3	4	5	6	7	8	9	10
Expiry	1	4.54	4.55	4.56	4.58	4.60	4.63	4.66	4.70	4.73	4.76
	2	4.55	4.57	4.59	4.62	4.65	4.69	4.73	4.76	4.79	...
	3	4.58	4.62	4.65	4.68	4.72	4.76	4.80	4.83	...	...
	4	4.65	4.68	4.72	4.76	4.80	4.84	4.87	...	...	...
	5	4.71	4.75	4.80	4.84	4.88	4.92	...	...	...	...
	6	4.80	4.84	4.89	4.93	4.97	...	...	...	...	...
	7	4.89	4.94	4.98	5.01	...	...	...	...	...	...
	8	4.99	5.03	5.06	...	...	...	...	...	...	...
	9	5.06	5.09	...	...	...	...	...	...	...	...
	10	5.13	...	...	...	...	...	...	...	...	...

Table 3.3: Initial swap rates (%) on October 17, 2007.

Here we compare the vegas of the Bermudan swaption under the HW and MR models. The mean reversion parameter a is fixed at 3% for both two driving processes so that the Bermudan prices produced by the two models are close and also comparable to those that are produced by the one step and weighted covariance models. We display the vegas in tables 3.4 and 3.5. The position of each vega corresponds to the implied volatility in the swaption matrix. Recall that the off reverse diagonal entries are all zero since the Bermudan swaption's price is not sensitive to the corresponding implied volatility inputs here.

Remark 2 *In practice, traders usually quote vega as the change in price when implied volatility increases by* 100 *basis points (bp) or* 1% *so we will scale the "true" vega by a factor of* 0.01*, i.e.* $\nu_{i,k} \to 0.01\nu_{i,k}$*. For example, the entry* 4.09 *in the first row and the last column of table 3.4 means that when* $\widetilde{\sigma}_{1,10}$ *increases by* 1% *the Bermudan price (with notional* 100 *million) will increase by* $40,900$*.*

Tenor		1	2	3	4	5	6	7	8	9	10
Expiry	1	0	0	0	0	0	0	0	0	0	4.09
	2	0	0	0	0	0	0	0	0	10.00	...
	3	0	0	0	0	0	0	0	9.48	...	...
	4	0	0	0	0	0	0	7.76	...	...	...
	5	0	0	0	0	0	6.23	...	...	...	...
	6	0	0	0	0	4.69	...	...	...	...	...
	7	0	0	0	3.74	...	...	...	...	...	...
	8	0	0	2.94	...	...	...	...	...	...	...
	9	0	2.16	...	...	...	...	...	...	...	...
	10	1.56	...	...	...	...	...	...	...	...	...

Table 3.4: The Bermudan swaption's scaled vegas (in 10^4) under the HW model.

Tenor		1	2	3	4	5	6	7	8	9	10
Expiry	1	0	0	0	0	0	0	0	0	0	6.28
	2	0	0	0	0	0	0	0	0	14.54	...
	3	0	0	0	0	0	0	0	12.12	...	...
	4	0	0	0	0	0	0	8.48	...	...	...
	5	0	0	0	0	0	5.62	...	...	...	...
	6	0	0	0	0	3.37	...	...	...	...	...
	7	0	0	0	1.86	...	...	...	...	...	...
	8	0	0	0.83	...	...	...	...	...	...	...
	9	0	0.22	...	...	...	...	...	...	...	...
	10	-0.02	...	...	...	...	...	...	...	...	...

Table 3.5: The Bermudan swaption's scaled vegas (in 10^4) under the MR model.

It is seen in figure 3.3 that the vegas for both models as a function of expiry display "humped" shapes whose peaks are attained at the same exercise date. We also observe that the HW vegas are lower than the MR vegas at the early exercise dates but higher at the later ones. Recall from Subsection 3.4.1 that this difference in vegas is actually caused by the difference in the correlation structures. This is rather important for a strongly correlation-dependent product like the Bermudan swaption. In the following, we will give a crude explanation on how a change in one of the co-terminal implied volatilities can affect the correlations of the co-terminal swap rates under the two models which will then clearly indicate their vegas.

Parametrization by expiry (MR): We recall that the MR process is unaltered

by bumping any co-terminal implied volatility. This then implies that the correlation structure of the model/co-terminal forward swap rates is unaltered, i.e. $\text{Corr}^{\text{mo}}(\ln y_{T_i}^{i,n+1-i}, \ln y_{T_j}^{j,n+1-j})$ which are approximately $\sqrt{\frac{\xi_{\min(T_i,T_j)}}{\xi_{\max(T_i,T_j)}}}$ are unchanged for all $i, j \leq n$.

Parametrization by time (HW): Bumping $\widetilde{\sigma}_{i,n+1-i}$ has an immediate effect on the HW process and hence the correlation structure of the model/co-terminal forward swap rates. Specifically, $\text{Corr}^{\text{mo}}(\ln y_{T_i}^{i,n+1-i}, \ln y_{T_j}^{j,n+1-j}) \approx \sqrt{\frac{\xi_{\min(T_i,T_j)}}{\xi_{\max(T_i,T_j)}}}$ will increase for $j > i$ but decrease for $j < i$. Heuristically, we have the following overall effect: on average the co-terminal forward swap rates will tend to be more correlated if i is small and less correlated if i is large.

We employ the following heuristic argument for the correlation's effects on the Bermudan price. The optionality of a Bermudan implies that the lower the correlations of the co-terminal swap rates get, the higher the price of the Bermudan swaption becomes. For more details of how correlations affect the Bermudan price, see Andersen and Piterbarg [2010] and Rebonato [2004] for example for reference. Hence, we draw the following conclusion on the vegas. If x is parameterized by time and i is small (early exercise date), the co-terminal forward swap rates will tend to be more correlated on average. This effect will cause the HW price to increase less than the MR price and make $\nu_{i,n+1-i}^{\text{hw}}$ lower than $\nu_{i,n+1-i}^{\text{mr}}$. On the other hand, if i is large (late exercise date), the co-terminal forward swap rates will tend to be less correlated on average. This then causes $\nu_{i,n+1-i}^{\text{hw}}$ to be higher than $\nu_{i,n+1-i}^{\text{mr}}$. This fundamental difference is the key observation which leads to very different hedging profiles as we shall see later.

Remark 3 *The MR vegas become very small or even negative at the end of the option which is possible under some circumstances in practice. See Appendix B in Pietersz and Pelsser [2004] for an explanation of negative vega for a two stock Bermudan option example.*

3.4.3 The Bermudan swaption's vegas under the one step and weighted covariance models

We test the one step and weighted covariance processes with different values of α and display their vega matrices in tables 3.6, 3.7, 3.8 and 3.9.

We first look at the vegas for the one step covariance model in table 3.6. The first thing to notice from this table is that the vega response starts shifting away from the reverse diagonal entries. We obtain a vega profile which assigns a significant contribution to the first column of the swaption matrix. The other vega entries are seen to be much smaller and very close to zero except for the co-terminal

ones. The vega behaviour of the first column can be seen from the local fit in Appendix 3.A.2. In this local fit step, apart from the reverse diagonal entries we see that shifting the first column has the most distinctive effect on the one step covariance terms that we estimate from the market. The one step covariance model is set up such that it responds to the changes in the one step covariances only (not the other covariances as considered in the weighted covariance model). Thus it is a shift in the first column or a reverse diagonal entry of the swaption matrix that has the largest vega response. Note that the swaptions corresponding to the first column are fairly illiquid so it would be desirable to use more implied volatilities to moderate the response to any inaccurate market signals. The results produced by the weighted covariance model indeed have this feature.

For the weighted covariance model, we observe different patterns for the vega response depending on different values of α. For example, when $\alpha = 0.05$ we observe a bigger response in the central part of the table compared with that when we use a much higher value of α, for instance $\alpha = 5$. For some particular rows, it is seen that those central entries even dominate the reverse diagonal and the first column. For larger values of α such as 0.3 in table 3.8, we observe a clearer trend in the vega entries. They tend to increase in tenor for each expiry. When α gets much higher (table 3.9), it is clear that the entries look very similar to the one step covariance case where the vegas from the central part become much more insignificant and dominated by the reverse diagonal and the first column. This is predictable as one step covariance is a special case of the weighted covariance model when α is very large.

Remark 4 *We get a few negative vega entries when $\alpha = 0.05$. Note that those in the reverse diagonal (the first two rows) are quite large in magnitude. One reason for this behaviour is the following. When we shift the co-terminal implied volatility $\widetilde{\sigma}_{i,n+1-i}$, the approximations from the market for the correlations $Corr^{ma}(\ln y_{T_i}^{i,n+1-i}, \ln y_{T_j}^{j,n+1-j})$ tend to increase for $j > i$. For a very low value of α, the model will take into account all these increases in correlations with high levels of impact on the Bermudan price since the geometric weight $p^{T_j - T_i}$ decays very slowly in $T_j - T_i$. Therefore, when i is small the overall increase in correlations of the co-terminal swap rates could be large which in turn leads to a decrease in price and hence negative vegas.*

Tenor		1	2	3	4	5	6	7	8	9	10
Expiry	1	0.27	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.60
	2	0.94	0.01	0.01	0.01	0.01	0.01	0.01	0.01	8.86	...
	3	1.46	0.01	0.01	0.01	0.01	0.01	0.00	7.79	...	...
	4	1.82	0.01	0.01	0.01	0.01	0.01	5.71	...	...	...
	5	2.01	0.02	0.02	0.02	0.01	4.02	...	...	...	...
	6	2.08	0.02	0.02	0.01	2.45	...	...	...	...	...
	7	2.05	0.02	0.01	1.56	...	...	...	...	...	...
	8	1.89	0.02	1.00	...	...	...	...	...	...	...
	9	1.68	0.43	...	...	...	...	...	...	...	...
	10	1.57	...	...	...	...	...	...	...	...	...

Table 3.6: The Bermudan swaption's scaled vegas (in 10^4) under the one step covariance model.

Tenor		1	2	3	4	5	6	7	8	9	10
Expiry	1	0.08	0.05	0.16	0.30	0.16	0.62	0.70	0.57	2.48	−1.05
	2	0.10	0.46	0.36	0.56	0.86	1.32	2.21	4.79	−1.06	...
	3	0.10	0.22	0.36	0.58	0.93	1.70	3.70	1.42	...	...
	4	0.06	0.16	0.29	0.54	0.95	2.15	3.49	...	...	...
	5	0.05	0.16	0.30	0.55	1.34	3.82	...	...	...	...
	6	0.02	0.10	0.16	0.46	3.89	...	...	...	...	...
	7	0.03	0.01	0.08	3.67	...	...	...	...	...	...
	8	−0.07	−0.24	3.25	...	...	...	...	...	...	...
	9	−0.18	2.25	...	...	...	...	...	...	...	...
	10	1.54	...	...	...	...	...	...	...	...	...

Table 3.7: The Bermudan swaption's scaled vegas (in 10^4) under the weighted covariance model ($\alpha = 0.05$).

Tenor		1	2	3	4	5	6	7	8	9	10
Expiry	1	0.10	0.13	0.18	0.25	0.15	0.35	0.30	0.38	0.65	1.86
	2	0.23	0.36	0.46	0.59	0.66	0.82	1.05	1.73	3.83	...
	3	0.22	0.41	0.53	0.67	0.81	1.14	1.87	3.75	...	...
	4	0.23	0.38	0.51	0.69	0.95	1.67	3.45	...	...	...
	5	0.18	0.35	0.53	0.81	1.48	3.00	...	...	...	...
	6	0.17	0.34	0.53	1.05	2.67	...	...	...	...	...
	7	0.13	0.31	0.71	2.68	...	...	...	...	...	...
	8	0.08	0.26	2.62	...	...	...	...	...	...	...
	9	0.08	2.03	...	...	...	...	...	...	...	...
	10	1.55	...	...	...	...	...	...	...	...	...

Table 3.8: The Bermudan swaption's scaled vegas (in 10^4) under the weighted covariance model ($\alpha = 0.3$).

Tenor		1	2	3	4	5	6	7	8	9	10
Expiry	1	0.25	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	3.59
	2	0.94	0.03	0.01	0.01	0.01	0.01	0.01	0.01	8.87	...
	3	1.44	0.03	0.01	0.01	0.01	0.00	0.00	7.77	...	...
	4	1.78	0.05	0.00	0.00	0.00	0.01	5.69	...	...	...
	5	1.98	0.05	0.00	0.00	0.00	4.03	...	...	...	...
	6	2.03	0.04	0.03	0.00	2.42	...	...	...	...	...
	7	2.03	0.08	0.02	1.58	...	...	...	...	...	...
	8	1.82	0.08	0.99	...	...	...	...	...	...	...
	9	1.66	0.46	...	...	...	...	...	...	...	...
	10	1.57	...	...	...	...	...	...	...	...	...

Table 3.9: The Bermudan swaption's scaled vegas (in 10^4) under the weighted covariance model ($\alpha = 5$).

3.4.4 The net market vegas for different parameterizations

We recall that the HW process is an example of parametrization by time and it has a certain vega profile with the responses only on the reverse diagonal. The one step and weighted covariance models move the vega response away from the reverse diagonal and this causes their hedging behaviours to be quite different from that of the HW model. However, their vega profiles are still very closely connected as they are different examples of parametrization by time. For the one step and weighted covariance models, we plot the sum of the vegas for each row (expiry) of the swaption

matrix. With our initial data, we observe that each row sum is roughly a constant that is independent of α and very close to the co-terminal vega on the same row of the HW model (figure 3.3).

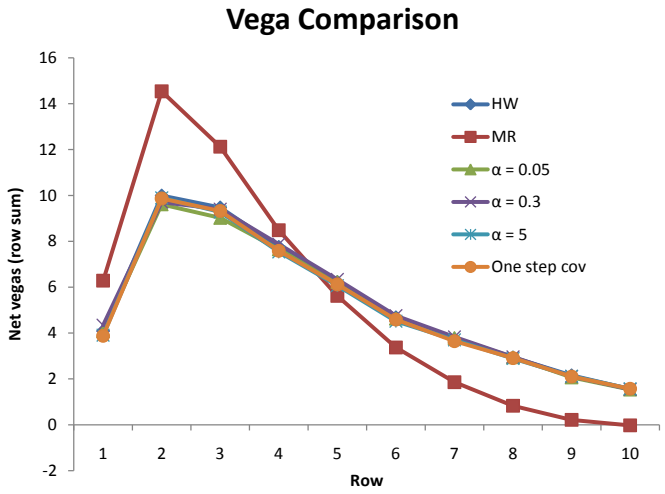


Figure 3.3: The (net) row sum of the scaled vegas (in 10^4) of a 11-years annual Bermudan swaption for different models and parameters.

We state this observation as a result.

Result 1 *For each $i = 1, \ldots, n$ and all $\alpha > 0$, under the assumptions that implied volatilities of the same expiry are not so variant with respect to tenor and the variances $\xi_{T_1}, \ldots, \xi_{T_n}$ are comparable between models the following relation holds true*

$$\sum_{k=1}^{n+1-i} \nu_{i,k}^{\alpha} \approx \nu_{i,n+1-i}^{hw} \tag{3.21}$$

In order to prove this result, we need the following sub-result that works with the log-transformation of the implied volatilities.

For $k = 1, \ldots, n+1-i$, let $\Sigma_{i,k} := \ln \widetilde{\sigma}_{i,k}$. We define the total derivative of the Bermudan swaption's price with respect to these log-implied volatilities as

$$\hat{\nu}_{i,k} := \frac{\mathrm{d}\hat{V}_{T_0}}{\mathrm{d}\Sigma_{i,k}}.$$

Again, in order to distinguish different models we denote $\hat{\nu}_{i,k}^{\alpha}$ for the one step and weighted covariance models. For the HW model, the equivalent co-terminal term is denoted by $\hat{\nu}_{i,n+1-i}^{\mathrm{hw}}$ and note that it is the only term that matters. We state the sub-result as a lemma.

Lemma 1 *For each $i = 1, \ldots, n$ and all $\alpha > 0$, under the second assumption in result 1 the following relation holds true*

$$\sum_{k=1}^{n+1-i} \hat{\nu}_{i,k}^{\alpha} \approx \hat{\nu}_{i,n+1-i}^{hw} \tag{3.22}$$

Proof: We will first prove that for the one step and weighted covariance models the row sum $\sum_{k=1}^{n+1-i} \hat{\nu}_{i,k}^{\alpha}$ is independent of α. Note that $\sum_{k=1}^{n+1-i} \hat{\nu}_{i,k}^{\alpha}$ roughly represents the effect of a parallel additive shift of $\Sigma_{i,k}$ for $k = 1, \ldots, n+1-i$ on the Bermudan price as we can see from the Taylor expansion of the price function $\hat{V}_{T_0}$. This is equivalent to a parallel multiplicative shift of $\widetilde{\sigma}_{i,k}$ for $k = 1, \ldots, n+1-i$ since

$$\Sigma_{i,k} \longrightarrow \Sigma_{i,k} + \ln \epsilon \iff \widetilde{\sigma}_{i,k} \longrightarrow \epsilon \widetilde{\sigma}_{i,k}.$$

One then can write

$$\sum_{k=1}^{n+1-i} \hat{\nu}_{i,k}^{\alpha} \approx \frac{\hat{V}_{T_0}(\epsilon\widetilde{\sigma}_{i,1}, \ldots, \epsilon\widetilde{\sigma}_{i,n+1-i}) - \hat{V}_{T_0}(\widetilde{\sigma}_{i,1}, \ldots, \widetilde{\sigma}_{i,n+1-i})}{\ln \epsilon},$$

where $\epsilon > 1$ and sufficiently small. It remains to show that the effect of the parallel multiplicative shift of the i$^{\text{th}}$ row of the swaption matrix on the Bermudan price is independent of α.

In the following we use the analysis obtained in Appendix 3.A. The main purpose is to assess the effects of the above parallel multiplicative shift on each of the estimated covariances that feed into the one step and weighted covariance models, i.e. $\text{Cov}^{\text{ma}}(\ln y_{T_i}^{i,n+1-i}, \ln y_{T_i}^{j,n+1-j})$ for $j = i+1, \ldots, n$. This can be seen via the effects on the covariances of the log-LIBORs. From the local fit in Appendix 3.A.2, we have the following approximation for the covariances of the log-LIBORs at each exercise date T_i

$$\widetilde{\sigma}_{i,k}^2 T_i \approx \sum_{l=i}^{i+k-1} \sum_{l^*=i}^{i+k-1} \zeta_l^{i,k}(0) \zeta_{l^*}^{i,k}(0) \text{Cov}^{\text{ma}}(\ln L_{T_i}^{l}, \ln L_{T_i}^{l^*}), \qquad k = 1, \ldots, n+1-i,$$

where $\{\zeta_l^{i,k}(0)\}_{l=i,\ldots,n}$ are constants that only depend on the initial discount curve (see Appendix 3.A). Therefore, under a parallel multiplicative shift of the i$^{\text{th}}$ row of the swaption matrix we have that

$$\text{Cov}^{\text{ma}}(\ln L_{T_i}^{l}, \ln L_{T_i}^{l^*}) \longrightarrow \epsilon^2 \text{Cov}^{\text{ma}}(\ln L_{T_i}^{l}, \ln L_{T_i}^{l^*}), \qquad l, l^* = i, \ldots, n.$$

Since the covariance of the log-swap rates can be approximated by summing up the covariances of the corresponding spanning log-LIBORs (see Appendix 3.A.2), we

have

$$\begin{aligned}
\mathrm{Cov}^{\mathrm{ma}}(\ln y_{T_i}^{i,n+1-i}, \ln y_{T_i}^{j,n+1-j}) \quad &\approx \quad \sum_{k=i}^{n}\sum_{l=j}^{n} \zeta_k^{i,n+1-i}(0)\zeta_l^{j,n+1-j}(0)\mathrm{Cov}^{\mathrm{ma}}(\ln L_{T_i}^k, \ln L_{T_i}^l) \\
\Rightarrow \mathrm{Cov}^{\mathrm{ma}}(\ln y_{T_i}^{i,n+1-i}, \ln y_{T_i}^{j,n+1-j}) \quad &\longrightarrow \quad \epsilon^2\mathrm{Cov}^{\mathrm{ma}}(\ln y_{T_i}^{i,n+1-i}, \ln y_{T_i}^{j,n+1-j}), \qquad j = i+1, \ldots, n.
\end{aligned}$$

Recall the market quantity $B_i = \sum_{j=i+1}^{n} p^{T_j - T_i}\mathrm{Cov}^{\mathrm{ma}}(\ln y_{T_i}^{i,n+1-i}, \ln y_{T_i}^{j,n+1-j})$. For all values of α, we then have that

$$B_i \longrightarrow \epsilon^2 B_i.$$

We further recall the construction of $\sqrt{\xi_{T_i}^{\alpha}}$ in equation (3.14) where we have

$$\sqrt{\xi_{T_i}^{\alpha}} = \frac{B_i}{\widetilde{\sigma}_{i,n+1-i}\sqrt{T_i}\sum_{j=i+1}^{n} p^{T_j - T_i}\sqrt{\frac{1}{\xi_{T_j}^{\alpha}}}\widetilde{\sigma}_{j,n+1-j}\sqrt{T_j}}.$$

Similar to the arguments in Section 3.4.1, the cascade structure in equation (3.15) implies that $\{\sqrt{\xi_{T_j}^{\alpha}}\}_{j=i+1,\ldots,n}$ are invariant under a parallel multiplicative shift of the i$^{\text{th}}$ row. It then follows from (3.14) that

$$\sqrt{\xi_{T_i}^{\alpha}} \longrightarrow \epsilon\sqrt{\xi_{T_i}^{\alpha}},$$

and hence further $\sqrt{\frac{1}{\xi_{T_i}^{\alpha}}}\widetilde{\sigma}_{i,n+1-i}\sqrt{T_i}$ is invariant. For $1 \leq s < i$, we have that

$$\sqrt{\xi_{T_s}^{\alpha}} = \frac{B_s}{\widetilde{\sigma}_{s,n+1-s}\sqrt{T_s}\sum_{j=s+1}^{n} p^{T_j - T_s}\sqrt{\frac{1}{\xi_{T_j}^{\alpha}}}\widetilde{\sigma}_{j,n+1-j}\sqrt{T_j}}.$$

When $s = i-1$, it is clear that $\sqrt{\xi_{T_s}^{\alpha}}$ is also invariant because $\sqrt{\frac{1}{\xi_{T_i}^{\alpha}}}\widetilde{\sigma}_{i,n+1-i}\sqrt{T_i}$ is invariant. Inductively, we have that $\{\sqrt{\xi_{T_s}^{\alpha}}\}_{1\leq s<i}$ are all invariant. It is now clear that a parallel multiplicative shift of the i$^{\text{th}}$ row will only shift $\sqrt{\xi_{T_i}^{\alpha}} \longrightarrow \epsilon\sqrt{\xi_{T_i}^{\alpha}}$ regardless of α.

Since $\Sigma_{i,k}$ is just the log of $\widetilde{\sigma}_{i,k}$ for each $k = 1, \ldots, n+1-i$, one can write the analogous formulae for $\hat{\nu}_{i,k}^{\alpha}$ following exactly the same arguments as in (3.18) and (3.19). Therefore, for the off reverse diagonal entries: $k = 1, \ldots, n-i$

$$\hat{\nu}_{i,k}^{\alpha} = \sum_{s=1}^{i} \frac{\partial \hat{V}_{T_0}}{\partial\sqrt{\xi_{T_s}^{\alpha}}} \times \frac{\mathrm{d}\sqrt{\xi_{T_s}^{\alpha}}}{\mathrm{d}\Sigma_{i,k}},$$

and for the reverse diagonal entry $k = n + 1 - i$

$$\hat{\nu}^{\alpha}_{i,n+1-i} = \frac{\partial \hat{V}_{T_0}}{\partial \Sigma_{i,n+1-i}} + \sum_{s=1}^{i} \frac{\partial \hat{V}_{T_0}}{\partial \sqrt{\xi^{\alpha}_{T_s}}} \times \frac{\mathrm{d}\sqrt{\xi^{\alpha}_{T_s}}}{\mathrm{d}\Sigma_{i,n+1-i}}.$$

Hence, we have the row sum

$$\begin{aligned} \sum_{k=1}^{n+1-i} \hat{\nu}^{\alpha}_{i,k} &= \frac{\partial \hat{V}_{T_0}}{\partial \Sigma_{i,n+1-i}} + \sum_{k=1}^{n+1-i} \sum_{s=1}^{i} \frac{\partial \hat{V}_{T_0}}{\partial \sqrt{\xi^{\alpha}_{T_s}}} \times \frac{\mathrm{d}\sqrt{\xi^{\alpha}_{T_s}}}{\mathrm{d}\Sigma_{i,k}} \\ &= \frac{\partial \hat{V}_{T_0}}{\partial \Sigma_{i,n+1-i}} + \sum_{s=1}^{i} \frac{\partial \hat{V}_{T_0}}{\partial \sqrt{\xi^{\alpha}_{T_s}}} \times \left(\sum_{k=1}^{n+1-i} \frac{\mathrm{d}\sqrt{\xi^{\alpha}_{T_s}}}{\mathrm{d}\Sigma_{i,k}} \right). \end{aligned}$$

Similar to that we discussed earlier, the sum $\sum_{k=1}^{n+1-i} \frac{\mathrm{d}\sqrt{\xi^{\alpha}_{T_s}}}{\mathrm{d}\Sigma_{i,k}}$ roughly represents the effect of a parallel multiplicative shift of the i$^{\text{th}}$ row of the swaption matrix on $\sqrt{\xi^{\alpha}_{T_s}}$ by looking at the Taylor expansion on $\sqrt{\xi^{\alpha}_{T_s}}$. We previously concluded for the one step and weighted covariance models that this shift will leave $\sqrt{\xi^{\alpha}_{T_j}}$ unchanged for $j \neq i$. It follows that $\sum_{k=1}^{n+1-i} \frac{\mathrm{d}\sqrt{\xi^{\alpha}_{T_s}}}{\mathrm{d}\Sigma_{i,k}} \approx 0$ for $s < i$. Hence

$$\sum_{k=1}^{n+1-i} \hat{\nu}^{\alpha}_{i,k} \approx \frac{\partial \hat{V}_{T_0}}{\partial \Sigma_{i,n+1-i}} + \frac{\partial \hat{V}_{T_0}}{\partial \sqrt{\xi^{\alpha}_{T_i}}} \times \sum_{k=1}^{n+1-i} \frac{\mathrm{d}\sqrt{\xi^{\alpha}_{T_i}}}{\mathrm{d}\Sigma_{i,k}}.$$

Since the parallel multiplicative shift of the i$^{\text{th}}$ row will alter $\sqrt{\xi^{\alpha}_{T_i}} \longrightarrow \epsilon\sqrt{\xi^{\alpha}_{T_i}}$, we then can write that

$$\sum_{k=1}^{n+1-i} \hat{\nu}^{\alpha}_{i,k} \approx \frac{\partial \hat{V}_{T_0}}{\partial \Sigma_{i,n+1-i}} + \frac{\partial \hat{V}_{T_0}}{\partial \sqrt{\xi^{\alpha}_{T_i}}} \times \frac{\epsilon\sqrt{\xi^{\alpha}_{T_i}} - \sqrt{\xi^{\alpha}_{T_i}}}{\ln \epsilon}. \tag{3.23}$$

Because we assume the $\sqrt{\xi^{\alpha}_{T_i}}$'s are similar for all α under our initial data, we can now conclude that $\sum_{k=1}^{n+1-i} \hat{\nu}^{\alpha}_{i,k}$ is independent of α.

We now prove the second part of the lemma that connects the weighted covariance process with the HW process. For the HW process, we recall that

$$\xi_{T_i} = \left(\frac{T_{n+1} - T_i}{(1 + \alpha_i y_0^{i,n+1-i})(\psi_{T_{n+1}} - \psi_{T_i})} \right)^2 \widetilde{\sigma}^2_{i,n+1-i} T_i,$$

where $\psi_{T_i} = \frac{1}{a}(1 - e^{-aT_i}), a > 0$. We fixed a at 3% so that the ξ_{T_i}'s are similar to the $\xi^{\alpha}_{T_i}$'s of the weighted covariance process. It is clear that shifting $\Sigma_{i,n+1-i} \longrightarrow$

$\Sigma_{i,n+1-i} + \ln\epsilon$ or equivalently $\widetilde{\sigma}_{i,n+1-i} \longrightarrow \epsilon\widetilde{\sigma}_{i,n+1-i}$ will only shift

$$\sqrt{\xi_{T_i}} \longrightarrow \epsilon\sqrt{\xi_{T_i}}, \tag{3.24}$$

and leave $\{\sqrt{\xi_{T_j}}\}_{j\neq i}$ remain unchanged. This is exactly the same effect that a parallel multiplicative shift of the i$^{\text{th}}$ row has on the weighted covariance process. We now write the analogous formula for $\hat{\nu}^{\text{hw}}_{i,n+1-i}$ following the same argument as in (3.20)

$$\hat{\nu}^{\text{hw}}_{i,n+1-i} = \frac{\partial \hat{V}_{T_0}}{\partial \Sigma_{i,n+1-i}} + \frac{\partial \hat{V}_{T_0}}{\partial \sqrt{\xi_{T_i}}} \times \frac{\mathrm{d}\sqrt{\xi_{T_i}}}{\mathrm{d}\Sigma_{i,n+1-i}}.$$

This then follows immediately from (3.24) that

$$\hat{\nu}^{\text{hw}}_{i,n+1-i} \approx \frac{\partial \hat{V}_{T_0}}{\partial \Sigma_{i,n+1-i}} + \frac{\partial \hat{V}_{T_0}}{\partial \sqrt{\xi_{T_i}}} \times \frac{\epsilon\sqrt{\xi_{T_i}} - \sqrt{\xi_{T_i}}}{\ln\epsilon},$$

which is approximately the same as (3.23) since ξ_{T_i} and $\xi^{\alpha}_{T_i}$ are comparable. The proof is now complete.

We are now able to prove result 1.

Proof of result 1: since $\Sigma_{i,k} = \ln\widetilde{\sigma}_{i,k}$, one can write the following for the one step and weighted covariance models

$$\nu^{\alpha}_{i,k} = \hat{\nu}^{\alpha}_{i,k}\frac{\mathrm{d}\Sigma_{i,k}}{\mathrm{d}\widetilde{\sigma}_{i,k}} = \hat{\nu}^{\alpha}_{i,k}\frac{1}{\widetilde{\sigma}_{i,k}}.$$

Hence

$$\sum_{k=1}^{n+1-i} \nu^{\alpha}_{i,k} = \sum_{k=1}^{n+1-i} \hat{\nu}^{\alpha}_{i,k}\frac{1}{\widetilde{\sigma}_{i,k}}.$$

For the HW model, the equivalent quantity is $\nu^{\text{hw}}_{i,n+1-i} = \hat{\nu}^{\text{hw}}_{i,n+1-i}\frac{1}{\widetilde{\sigma}_{i,n+1-i}}$. Observe that if the implied volatility is not so variant in tenor k, we will be able to remove $\widetilde{\sigma}^{-1}_{i,k}$ from the sum and replace them by a constant C. This assumption is also supported by the data we work with in this chapter (table 3.2). It then follows by lemma 1 that

$$\sum_{k=1}^{n+1-i} \nu^{\alpha}_{i,k} \approx C\sum_{k=1}^{n+1-i} \hat{\nu}^{\alpha}_{i,k} \approx C\hat{\nu}^{\text{hw}}_{i,n+1-i} \approx \nu^{\text{hw}}_{i,n+1-i}.$$

The proof is now complete.

3.5 A hedging result

In this section, we construct a vega-delta neutral portfolio for a Bermudan swaption under different models. The portfolio will consist of vanilla swaptions and interest

rate swaps. We then calculate the gamma profile of the portfolio for each model and compare them accordingly. Here we employ the same example of Bermudan swaption as we considered in Section 3.4 with the data provided in tables 3.2 and 3.3.

We first look at the gamma profiles of the HW and MR hedging portfolios in Section 3.5.1. We then continue our investigation with the one step and weighted covariance models in Section 3.5.2. In all cases, we want to stress that the vegas of the Bermudan are the main factor that affects the gamma profile of the hedging portfolio. For the first two models, we observe a distinct difference between them in terms of the total gamma of vega-delta neutral portfolio. These will be addressed in detail in Section 3.5.1.3. It is also seen in this section that the parallel gamma profiles of these models are qualitatively similar and not ideal in practice. Due to the more realistic specification of the later models, one can actually hope that the gamma profile of the hedging portfolio will be improved. This is a potential advantage of the one step and weighted covariance models and we will analyze it in detail in Section 3.5.2.3.

3.5.1 A hedging portfolio for the HW and MR models

We now proceed to the construction of the vega-delta neutral portfolios for the HW and MR models.

3.5.1.1 Vega hedge

We first construct a vega neutral portfolio. Since the Bermudan prices under the HW and MR models are only sensitive to the changes in the co-terminal implied volatilities, we will only need to hedge those risks by trading suitable proportions of the co-terminal vanilla swaptions. At this stage, the portfolio will consist of a Bermudan swaption with today's value $\hat{V}_{T_0}$ and an appropriate proportion N_i^{sption} of the corresponding i$^{\text{th}}$ co-terminal vanilla swaption with today's value $\widetilde{V}_{T_0}^{i,n+1-i}$ for each $i = 1, \ldots, n$. We require the portfolio to be vega neutral, i.e. for each $i = 1, \ldots, n$, N_i^{sption} is chosen such that

$$\frac{\mathrm{d}\hat{V}_{T_0}}{\mathrm{d}\widetilde{\sigma}_{i,n+1-i}} + N_i^{\text{sption}} \frac{\mathrm{d}\widetilde{V}_{T_0}^{i,n+1-i}}{\mathrm{d}\widetilde{\sigma}_{i,n+1-i}} = \nu_{i,n+1-i} + N_i^{\text{sption}} \frac{\mathrm{d}\widetilde{V}_{T_0}^{i,n+1-i}}{\mathrm{d}\widetilde{\sigma}_{i,n+1-i}} = 0,$$

where the derivatives can be calculated numerically by the finite difference method and note that $\nu_{i,n+1-i}$ indicates the i$^{\text{th}}$ bucket vega of the Bermudan for the HW and MR models. Hence, we have that

$$N_i^{\text{sption}} = -\nu_{i,n+1-i} \Big/ \left(\frac{\mathrm{d}\widetilde{V}_{T_0}^{i,n+1-i}}{\mathrm{d}\widetilde{\sigma}_{i,n+1-i}} \right).$$

We display the vegas of the Bermudan calculated for the HW and MR models together in table 3.10 and the co-terminal vanilla swaptions' vegas (with notional being 1 million) in table 3.11. The calculations of N_i^{sption} then follow directly and we display the results in table 3.12.

i	1	2	3	4	5	6	7	8	9	10
HW	4.09	10.00	9.48	7.76	6.23	4.69	3.74	2.94	2.16	1.56
MR	6.28	14.54	12.12	8.48	5.62	3.37	1.86	0.83	0.22	-0.02

Table 3.10: The Bermudan swaption's scaled vegas (in 10^4) for the HW and MR models.

i	1	2	3	4	5	6	7	8	9	10
vanilla	0.27	0.35	0.38	0.38	0.36	0.32	0.27	0.21	0.15	0.08

Table 3.11: The co-terminal vanilla swaptions' scaled vegas (in 10^4).

i	1	2	3	4	5	6	7	8	9	10
HW	-15	-28	-25	-20	-17	-15	-14	-14	-15	-21
MR	-23	-41	-32	-22	-16	-11	-7	-4	-1	0

Table 3.12: Vega hedging (N_i^{sption}) for the HW and MR models.

In table 3.12, one observes that N_i^{sption}'s of the MR model are about one and a half times as big in magnitude as those of the HW model for small i. The gap between N_i^{sption}'s of the two models gets smaller as i increases and after a certain i we observe the reverse situation, i.e. N_i^{sption}'s of the MR model become much smaller in magnitude compared with those of the HW model. This is an immediate result from the difference in their vegas which is a direct consequence of the difference between the parameterizations by time and by expiry as we explored in Sections 3.3.2 and 3.4.2.

3.5.1.2 Delta hedge

After the vega hedging step, the hedging portfolio is no longer exposed to the vega risks but still exposed to the delta risks, i.e. the risks with respect to the movements of the underlyings. The next step is, therefore, to neutralize the delta risks but still maintain the vega neutrality. We will do so by using the co-initial swaps with today's values $V_{T_0}^{0,i}$ for $i = 1, \ldots, n+1$ as the hedging instruments. Note that

the corresponding swap rates of the co-initial swaps are $y^{0,i}$ which start today and mature at T_i for $i = 1, \ldots, n+1$. We will work with the co-initial swap rates as the underlyings instead of the pure discount bonds as their prices are directly available in the market.

We first define the deltas of a Bermudan to be

$$\hat{\Delta}_j := \frac{\mathrm{d}\hat{V}_{T_0}}{\mathrm{d}y^{0,j}},$$

for $j = 1, \ldots, n+1$. These derivatives can be calculated by the finite difference method. At this stage, the current vega neutral portfolio has non-zero deltas.

Because the co-initial swaps are not sensitive to any changes in the implied volatilities, adding them to the portfolio will not affect the vega neutrality. Denote the proportions of the co-initial swaps that we wish to acquire by N_i^{swap} for $i = 1, \ldots, n+1$. We now have a new hedging portfolio with today's value $V_{T_0}^{\text{port}}$

$$V_{T_0}^{\text{port}} = \hat{V}_{T_0} + \sum_{i=1}^{n} N_i^{\text{sption}} \widetilde{V}_{T_0}^{i,n+1-i} + \sum_{i=1}^{n+1} N_i^{\text{swap}} V_{T_0}^{0,i}.$$

The proportions $\{N_i^{\text{swap}}\}_{i=1,\ldots,n+1}$ need to be chosen so that the hedging portfolio becomes delta neutral. The j$^{\text{th}}$ delta position of the portfolio is denoted by Δ_j where

$$\Delta_j := \frac{\mathrm{d}V_{T_0}^{\text{port}}}{\mathrm{d}y^{0,j}},$$

for $j = 1, \ldots, n+1$. Therefore, in order to neutralize the delta risks we need to solve for vector $\mathbf{N}^{\text{swap}} = (N_1^{\text{swap}}, \ldots, N_{n+1}^{\text{swap}})$ so that $\Delta_j = 0$ for $j = 1, \ldots, n+1$. This is a straightforward task and it effectively requires a matrix multiplication to get the vector $\mathbf{N}^{\text{swap}}$. The portfolios for the HW and MR models are now vega-delta neutral.

We present the Bermudan's deltas and N_i^{swap}'s in figure 3.4 and table 3.13 respectively. Note that there is a large jump in the last delta of the Bermudan which is not straightforward at first glance. The reason is that shifting the last co-initial swap rate will shift the last LIBOR and result in a parallel shift of all the co-terminal swap rates. On the other hand, the equivalent effect of shifting the other co-initial swap rates is that it will only decrease slightly one of the co-terminal swap rates and leave the rest almost unchanged. As a result, the last delta is positive and large in magnitude while the others are small and negative. These effects can be seen via the one to one correspondence between the co-initial swap rates and the discount bonds (see Appendix 3.B).

Remark 5 *In practice, delta is usually quoted as the change in price when the*

underlying rate moves by 1 *bp* (0.0001) *so we will scale the "true" delta as calculated above by a factor of* 0.0001, *i.e.* $\hat{\Delta}_j \to 0.0001\hat{\Delta}_j$. *For example, a delta of around* -3000 *in figure 3.4 means that the Bermudan price (with notional* 100 *million) will decrease by around* 3000 *if the corresponding co-initial swap rate increases by* 1 *bp.*

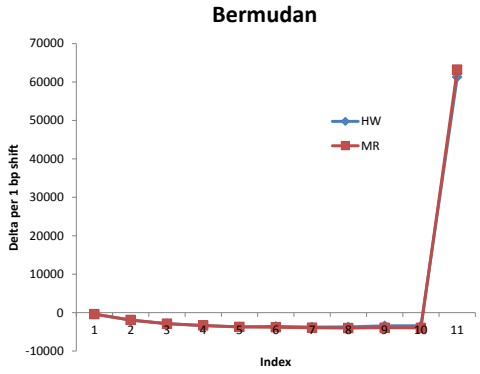


Figure 3.4: The scaled deltas of the Bermudan under the HW and MR models.

i	1	2	3	4	5	6	7	8	9	10	11
HW	-1.4	-2.0	-1.3	-0.9	-0.8	-1.2	-1.6	-2.6	-4.2	-8.2	21.8
MR	-4.4	-7.6	-4.7	-2.0	-0.0	1.4	2.5	3.5	4.2	4.8	-0.9

Table 3.13: Delta hedging (N_i^{swap}) for the HW and MR models (correspond to swaps with notional $N = 1$ million).

3.5.1.3 The gammas of the HW and MR hedging portfolios

Recall that each of the deltas is itself a function of the co-initial swap rates so they will move whenever these rates move. This change of deltas requires us to re-balance the portfolio since the previous portfolio is no longer delta-neutral. It is then important to consider the gammas of the portfolio as a measure of the sensitivity of the deltas. As the re-balancing cost could sometimes be high, it is desirable from a practical point of view that the gammas of the portfolio should have as small magnitudes as possible which implies that there is very little need for re-balancing the portfolio. For each $i = 1, \ldots, n+1$, we carry out a parallel shift of the co-initial swap rates with $\epsilon = 1$ bp (0.0001) and re-evaluate the delta to calculate the gamma. For the Bermudan, we then end up with the gamma vector $\hat{\mathbf{\Gamma}}$ where each of the

co-ordinates is

$$\hat{\Gamma}_i := \frac{\hat{\Delta}_i(y^{0,1}+\epsilon,\ldots,y^{0,n+1}+\epsilon)-\hat{\Delta}_i(y^{0,1},\ldots,y^{0,n+1})}{\epsilon}, \qquad i=1,\ldots,n+1. \tag{3.25}$$

For the vega-delta neutral portfolio, we have the corresponding gamma vector $\mathbf{\Gamma} = (\Gamma_1,\ldots,\Gamma_{n+1})$. The quantity in (3.25) is referred to as the parallel gamma and tells the trader how much the delta moves when the market moves. It is a proxy for the row sums of the full gamma matrix

$$\hat{\Gamma}_i \approx \sum_{j=1}^{n+1} \frac{\mathrm{d}^2\hat{V}_{T_0}}{\mathrm{d}y^{0,i}\mathrm{d}y^{0,j}}.$$

Remark 6 *Again, we note that although the "true" gammas are the second order derivatives of the price with respect to the underlyings, it is market convention to quote them as the change in the scaled deltas (by a factor of* 0.0001*) when the underlying rate(s) increases by* 1 *bp. For example, one will quote the parallel gamma* $\hat{\Gamma}_i$ *as calculated in* (3.25) *as*

$$0.0001\hat{\Delta}_i(y^{0,1}+\epsilon,\ldots,y^{0,n+1}+\epsilon)-0.0001\hat{\Delta}_i(y^{0,1},\ldots,y^{0,n+1}).$$

Therefore, the gammas we display later are the "true" gammas scaled by a factor of 10^{-8}.

Figure 3.5 illustrates the row sums of the gamma matrix and the gamma vector $\hat{\mathbf{\Gamma}}$ of the Bermudan under the HW model (similar for all other models). As one can expect, the two calculations show similar results with some small fluctuations from the proxy calculation (3.25). Note that the large jump in the last gamma is a direct consequence of the jump in the last delta.

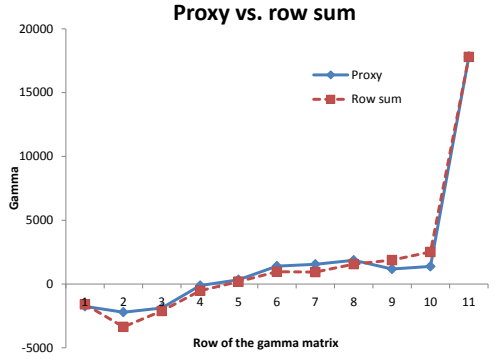


Figure 3.5: Proxy vs. row sum of the gamma matrix under the HW model.

An aggregated gamma quantity of interest is the total gamma. This is the sum of all entries in the gamma matrix $\Gamma = \sum_{i=1}^{n+1} \Gamma_i$ and it measures the total gamma exposure of the vega-delta neutral portfolio. We approximate this quantity by adding up all co-ordinates of the gamma vector since each co-ordinate is a proxy for the row sum of the gamma matrix. In table 3.14, we display the scaled total gamma for each component of the HW and MR portfolios. The result shows that the magnitude of the total gamma of the MR portfolio is around twice as big as that of the HW portfolio. This difference mainly comes from the big gap between the total gamma contributions of the co-terminal swaptions of the two portfolios. On the other hand, the contributions from the co-initial swaps are seen to be much smaller in magnitude so we will not discuss them in detail.

	Bermudan	vanilla swaptions	co-initial swaps	portfolio
HW	196	-286	2	-88
MR	192	-369	-7	-184

Table 3.14: Scaled total gammas of the HW and MR portfolios.

In order to understand the difference in the total gammas of the two portfolios, we plot the total gamma contribution from each individual co-terminal swaption in figure 3.6. Note that the total gamma contribution of the vanilla swaptions (-286 for HW and -369 for MR after scaling) is basically $\sum_{i=1}^{n} N_i^{\text{sption}} \sum_{j=1}^{n+1} \widetilde{\Gamma}_j^{i,n+1-i}$. The plot in figure 3.6 displays $N_i^{\text{sption}} \sum_{j=1}^{n+1} \widetilde{\Gamma}_j^{i,n+1-i}$ after scaling for each i. The difference in the negative peaks is clearly caused by the difference in N_i^{sption}'s as we observed in table 3.12. Recall that the difference in N_i^{sption}'s is a consequence of the difference in the vegas of the Bermudan for the two models which is characterized by the difference between the parametrizations by time and by expiry. From this analysis, one then concludes that the HW portfolio has a better total gamma than the MR portfolio. It indicates that in a wider context a parametrization by time process will potentially lead to a better total gamma profile of the hedging portfolio than a parametrization by expiry process.

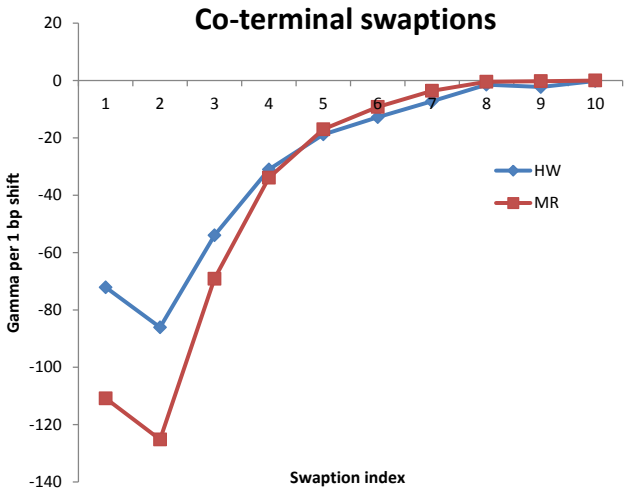


Figure 3.6: Scaled total gamma contributions from the co-terminal vanilla swaptions.

We now look at the gamma vectors of the HW and MR portfolios. Figure 3.7 displays the scaled gamma vectors of the portfolios before the hedge (just the Bermudan) and after the hedge (together with the co-terminal swaptions and the co-initial swaps). The results show that the gamma vectors of the Bermudan swaption for both models are very similar. After the hedge, the portfolios of the two models still have qualitatively very similar gamma profiles. Note that the large jump in the last gamma in the left plot is a direct consequence of the jump in the last delta in figure 3.4. From the right plot in figure 3.7, we see that for both models as all co-initial swap rates increase the deltas Δ_i's will increase for $i < n+1$ and decrease for $i = n+1$. We further observe that the rates of the increase are much lower than the rate of the decrease. Clearly, we do not have a good hedge of the last gamma for both the HW and MR models.

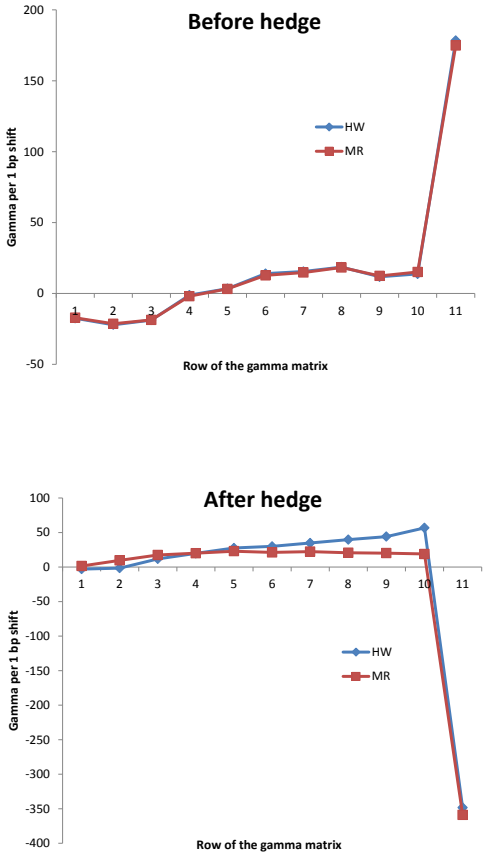


Figure 3.7: Scaled gamma vectors of the HW and MR portfolios before and after the hedge.

To gain more insight into the gamma behaviours, we will look at the gamma contribution from each hedging component. We further denote the gamma vector of the co-terminal swaption with today's value $\widetilde{V}_{T_0}^{i,n+1-i}$ by $\widetilde{\mathbf{\Gamma}}^{\mathbf{i,n+1-i}} = (\widetilde{\Gamma}_1^{i,n+1-i}, \ldots, \widetilde{\Gamma}_{n+1}^{i,n+1-i})$. Similarly, we denote the gamma vector of the co-initial swap with today's value $V_{T_0}^{0,i}$ by $\mathbf{\Gamma}^{\mathbf{0,i}} = (\Gamma_1^{0,i}, \ldots, \Gamma_{n+1}^{0,i})$. Recall that the portfolio consists of

$$V_{T_0}^{\text{port}} = \hat{V}_{T_0} + \sum_{i=1}^{n} N_i^{\text{sption}} \widetilde{V}_{T_0}^{i,n+1-i} + \sum_{i=1}^{n+1} N_i^{\text{swap}} V_{T_0}^{0,i},$$

and hence the portfolio's parallel gamma can be written as

$$\Gamma_j = \underbrace{\hat{\Gamma}_j}_{\text{Bermudan}} + \sum_{i=1}^{n} N_i^{\text{sption}} \underbrace{\widetilde{\Gamma}_j^{i,n+1-i}}_{\text{Co-terminal swaptions}} + \sum_{i=1}^{n+1} N_i^{\text{swap}} \underbrace{\Gamma_j^{0,i}}_{\text{Co-initial swaps}},$$

for $j = 1, \ldots, n+1$. Figure 3.8 represents the scaled gamma contribution vector from each hedging component of the portfolio. The left and right plots display $\sum_{i=1}^{n} N_i^{\text{sption}} \widetilde{\Gamma}_j^{i,n+1-i}$ and $\sum_{i=1}^{n+1} N_i^{\text{swap}} \Gamma_j^{0,i}$ after scaling respectively for each $j = 1, \ldots, n+1$. Note that the index i for both sums indicates the corresponding co-terminal swaption and the co-initial swap. While the co-initial swaps do not have very large gammas, the contribution from the co-terminal swaptions clearly determines the gamma behaviour of the whole portfolio. We observe in the left plot of figure 3.8 that for both models the last gamma contribution $\sum_{i=1}^{n} N_i^{\text{sption}} \widetilde{\Gamma}_{n+1}^{i,n+1-i}$ is extremely negative and it pulls the last gamma Γ_{n+1} of the portfolio down to be also very negative. This seems to be a common problem for both models. The reason is that both the HW and MR models assign the vega responses to only the reverse diagonal of the swaption matrix. This results in large values in magnitude of N_i^{sption}'s which then lead to very negative values of $\sum_{i=1}^{n} N_i^{\text{sption}} \widetilde{\Gamma}_{n+1}^{i,n+1-i}$. In Section 3.5.2.3, we will examine for the one step and weighted covariance models how moving the vega responses away from the reverse diagonal and spreading them over the swaption matrix can influence (improve) the gamma profile of the portfolio.

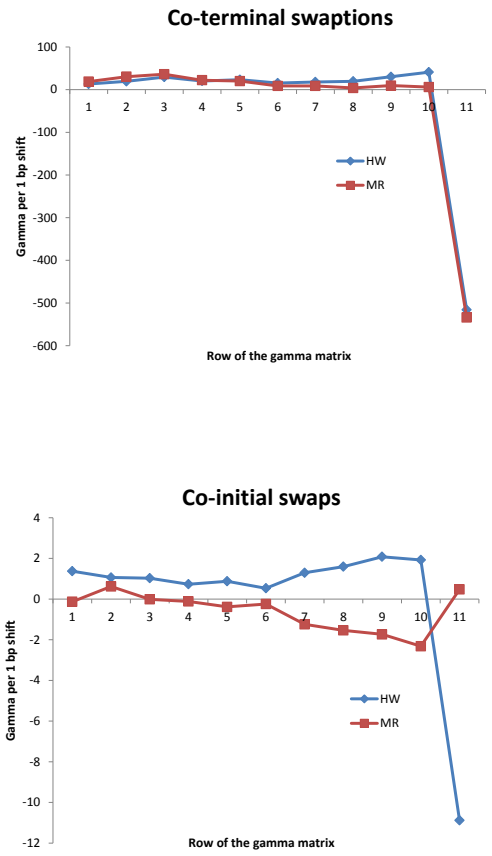


Figure 3.8: Scaled gamma contribution vectors of the co-terminal swaptions and the co-initial swaps for the HW and MR portfolios.

3.5.2 A hedging portfolio for the one step and weighted covariance models

This section discusses the vega-delta neutral portfolios for the one step and weighted covariance models. Similar to our treatment of the HW and MR models, we again use the vanilla swaptions and the co-initial swaps to vega-delta hedge the Bermudan.

3.5.2.1 Vega hedge

In Section 3.4.4, we discussed the vegas of a Bermudan computed for the one step and weighted covariance models. Since Bermudan prices under these models respond to the changes of certain implied volatilities in the swaption matrix, we will need to hedge those risks by trading suitable proportions of the appropriate vanilla swaptions. At this stage, the portfolio will consist of a Bermudan swaption with today's value $\hat{V}_{T_0}$ and an appropriate proportion $N_{i,j}^{\text{sption}}$ of the corresponding vanilla swaption with today's value $\widetilde{V}_{T_0}^{i,j}$ (with notional being 1 million) for each $i = 1, \ldots, n$

and $j = 1, \ldots, n+1-i$. We require the portfolio to be vega neutral, i.e. for each $i = 1, \ldots, n$ and $j = 1, \ldots, n+1-i$, $N_{i,j}^{\text{sption}}$ is chosen such that

$$\frac{\mathrm{d}\hat{V}_{T_0}}{\mathrm{d}\widetilde{\sigma}_{i,j}} + N_{i,j}^{\text{sption}} \frac{\mathrm{d}\widetilde{V}_{T_0}^{i,j}}{\mathrm{d}\widetilde{\sigma}_{i,j}} = \nu_{i,j}^{\alpha} + N_{i,j}^{\text{sption}} \frac{\mathrm{d}\widetilde{V}_{T_0}^{i,j}}{\mathrm{d}\widetilde{\sigma}_{i,j}} = 0,$$

where the derivatives can be calculated numerically by the finite difference method. The vegas of vanilla swaptions are displayed in table 3.15. We recall that $\nu_{i,j}^{\alpha}$ is the (i,j) entry in the Bermudan's vega matrix of the one step and weighted covariance models. Hence, we have that

$$N_{i,j}^{\text{sption}} = -\nu_{i,j}^{\alpha} / \left(\frac{\mathrm{d}\widetilde{V}_{T_0}^{i,j}}{\mathrm{d}\widetilde{\sigma}_{i,j}} \right).$$

We now present the results from the vega-hedging step in the following tables.

Tenor		1	2	3	4	5	6	7	8	9	10
Expiry	1	0.03	0.05	0.08	0.10	0.13	0.16	0.19	0.21	0.24	0.27
	2	0.04	0.08	0.12	0.16	0.20	0.24	0.28	0.32	0.35	...
	3	0.05	0.10	0.15	0.20	0.25	0.29	0.34	0.38	...	...
	4	0.06	0.12	0.17	0.23	0.28	0.33	0.38	...	...	...
	5	0.06	0.13	0.19	0.25	0.30	0.36	...	...	...	...
	6	0.07	0.13	0.20	0.26	0.32	...	...	...	...	...
	7	0.07	0.14	0.21	0.27	...	...	...	...	...	...
	8	0.07	0.14	0.21	...	...	...	...	...	...	...
	9	0.07	0.15	...	...	...	...	...	...	...	...
	10	0.08	...	...	...	...	...	...	...	...	...

Table 3.15: The vanilla swaptions' scaled vegas (in 10^4).

Tenor		1	2	3	4	5	6	7	8	9	10
Expiry	1	-10.0	-0.0	-0.0	-0.0	-0.0	-0.0	-0.0	-0.0	-0.0	-13.4
	2	-22.7	-0.1	-0.1	-0.1	-0.1	-0.0	-0.0	-0.0	-25.2	...
	3	-28.6	-0.1	-0.1	-0.1	-0.1	-0.0	-0.0	-20.5	...	...
	4	-31.1	-0.1	-0.1	-0.1	-0.1	-0.0	-15.1	...	...	...
	5	-31.5	-0.2	-0.1	-0.1	-0.0	-11.3	...	...	...	...
	6	-30.5	-0.2	-0.1	-0.0	-7.7	...	...	...	...	...
	7	-28.7	-0.2	-0.1	-5.8	...	...	...	...	...	...
	8	-25.6	-0.1	-4.7	...	...	...	...	...	...	...
	9	-22.4	-2.9	...	...	...	...	...	...	...	...
	10	-20.9	...	...	...	...	...	...	...	...	...

Table 3.16: Vega hedging ($N_{i,j}^{\text{sption}}$) for the one step covariance model.

Tenor		1	2	3	4	5	6	7	8	9	10
Expiry	1	-3.0	-1.0	-2.1	-2.8	-1.2	-3.9	-3.8	-2.7	-10.3	3.9
	2	-2.4	-5.7	-3.0	-3.5	-4.3	-5.6	-8.0	-15.2	3.0	...
	3	-1.9	-2.2	-2.4	-2.9	-3.8	-5.8	-11.0	-3.7	...	...
	4	-1.1	-1.4	-1.7	-2.4	-3.4	-6.5	-9.2	...	...	...
	5	-0.8	-1.3	-1.6	-2.3	-4.4	-10.7	...	...	...	...
	6	-0.3	-0.8	-0.8	-1.8	-12.2	...	...	...	...	...
	7	-0.4	-0.1	-0.4	-13.6	...	...	...	...	...	...
	8	0.9	1.7	-15.4	...	...	...	...	...	...	...
	9	2.4	-15.4	...	...	...	...	...	...	...	...
	10	-20.4	...	...	...	...	...	...	...	...	...

Table 3.17: Vega hedging ($N_{i,j}^{\text{sption}}$) for the weighted covariance model ($\alpha = 0.05$).

Tenor		1	2	3	4	5	6	7	8	9	10
Expiry	1	-3.8	-2.4	-2.3	-2.4	-1.2	-2.2	-1.6	-1.8	-2.7	-6.9
	2	-5.6	-4.4	-3.8	-3.7	-3.3	-3.4	-3.8	-5.5	-10.9	...
	3	-4.3	-4.1	-3.6	-3.4	-3.3	-3.9	-5.6	-9.9	...	...
	4	-4.0	-3.3	-3.0	-3.0	-3.4	-5.1	-9.1	...	...	...
	5	-2.8	-2.8	-2.8	-3.3	-4.9	-8.4	...	...	...	...
	6	-2.5	-2.5	-2.7	-4.0	-8.4	...	...	...	...	...
	7	-1.9	-2.2	-3.4	-9.9	...	...	...	...	...	...
	8	-1.1	-1.8	-12.3	...	...	...	...	...	...	...
	9	-1.0	-13.9	...	...	...	...	...	...	...	...
	10	-20.5	...	...	...	...	...	...	...	...	...

Table 3.18: Vega hedging ($N_{i,j}^{\text{sption}}$) for the weighted covariance model ($\alpha = 0.3$).

Tenor		1	2	3	4	5	6	7	8	9	10
Expiry	1	-9.2	-0.3	-0.2	-0.1	-0.1	-0.0	-0.0	-0.0	-0.0	-13.3
	2	-22.9	-0.4	-0.1	-0.1	-0.0	-0.0	-0.0	-0.0	-25.2	...
	3	-28.2	-0.3	-0.1	-0.1	-0.0	-0.0	-0.0	-20.4	...	...
	4	-30.4	-0.5	-0.0	-0.0	-0.0	-0.0	-15.0	...	...	...
	5	-30.9	-0.4	-0.0	-0.0	-0.0	-11.3	...	...	...	...
	6	-29.7	-0.3	-0.2	0.0	-7.6	...	...	...	...	...
	7	-28.4	-0.6	-0.1	-5.8	...	...	...	...	...	...
	8	-24.7	-0.6	-4.7	...	...	...	...	...	...	...
	9	-22.1	-3.2	...	...	...	...	...	...	...	...
	10	-20.9	...	...	...	...	...	...	...	...	...

Table 3.19: Vega hedging ($N_{i,j}^{\text{sption}}$) for the weighted covariance model ($\alpha = 5$).

3.5.2.2 Delta hedge

Similar to the HW and MR models, we use the co-initial swaps with today's values $V_{T_0}^{0,i}$ for $i = 1, \ldots, n+1$ to delta hedge the vega neutral portfolio. To be specific, we construct the hedging portfolio with today's value $V_{T_0}^{\text{port}}$

$$V_{T_0}^{\text{port}} = \hat{V}_{T_0} + \sum_{i=1}^{n} \sum_{j=1}^{n+1-i} N_{i,j}^{\text{sption}} \widetilde{V}_{T_0}^{i,j} + \sum_{i=1}^{n+1} N_i^{\text{swap}} V_{T_0}^{0,i}.$$

In order to neutralize the deltas, one then needs to solve for vector $\mathbf{N}^{\text{swap}} = (N_1^{\text{swap}}, \ldots, N_{n+1}^{\text{swap}})$ so that $\Delta_j = 0$ where each Δ_j is the overall j$^{\text{th}}$ delta posi-

tion of the portfolio for $j = 1, \ldots, n+1$. This step again requires us to effectively do a matrix multiplication to obtain the vector $\mathbf{N}^{\text{swap}}$.

Since the deltas of the Bermudan under the one step and weighted covariance models are very similar to those under the HW and MR models, there is no extra interest in plotting them. We display the vectors $\mathbf{N}^{\text{swap}}$ for the one step and weighted covariance models in table 3.20.

i	1	2	3	4	5	6	7	8	9	10	11
$\alpha = 0.05$	-3.2	-6.3	-3.5	0.1	0.9	1.8	3.8	6.0	5.8	11.5	-19.1
$\alpha = 0.3$	-4.0	-6.0	-3.6	-1.4	1.3	2.1	3.9	5.4	5.7	8.6	-13.8
$\alpha = 5$	-3.6	-6.2	-2.3	-0.1	1.7	2.9	3.2	4.4	4.3	5.5	-8.3
One step cov	-3.7	-5.9	-2.6	-0.3	1.6	2.6	3.6	4.3	4.6	5.4	-8.3

Table 3.20: Delta hedging (N_i^{swap}) for the one step and the weighted covariance models (correspond to swaps with notional $N = 1$ million).

3.5.2.3 The gammas of the one step and weighted covariance hedging portfolios

In this section, we carry out a similar comparison as in Section 3.5.1.3 but for the HW and the one step and weighted covariance portfolios. We first report the total gamma of each portfolio produced by the one step and weighted covariance models in table 3.21. Recall that the total gamma is approximated by summing up all coordinates of the parallel gamma vector. One can see that the results are very similar and comparable between models. Furthermore, they slightly deviate from that of the HW model and are still much smaller in magnitude compared with that of the MR model (see table 3.14). This again supports our findings in Section 3.5.1.3 that a parametrization by time process potentially leads to a better vega-delta neutral portfolio than a parametrization by expiry process in terms of the total gamma.

Note that the overall contributions from the vanilla swaptions in all four portfolios are quite close to the overall contributions from the co-terminal swaptions in the HW case. In order to calculate these contributions for the one step and weighted covariance models, one first calculates the total gamma for each vanilla swaption involved in the vega hedge and then multiplies with the corresponding proportion of holding $N_{i,j}^{\text{sption}}$ in the portfolio. We then sum these products up to obtain the overall contribution. It turns out that as we sum them up for each row of the swaption matrix, we effectively get the same plot as figure 3.6 of the HW model. This observation can be linked back to the vega "row sum" observation in result 1. The Bermudan's vega $\nu_{i,j}^{\alpha}$ of the one step and weighted covariance models is directly connected to the proportion $N_{i,j}^{\text{sption}}$ in the vega hedge. Intuitively, the similarity

between the vega row sum $\sum_{j=1}^{n+1-i} \nu_{i,j}^{\alpha}$ and the HW's i$^{\text{th}}$ bucket vega $\nu_{i,n+1-i}^{\text{hw}}$ for each $i = 1, \ldots, n$ should also carry over to the row sum[2] of the total gamma to some extent due to the direct connection of the proportions of vanilla swaptions in their portfolios, i.e. $\{N_{i,j}^{\text{sption}}\}_{j=1,\ldots,n+1-i}$ of the one step and weighted covariance models and N_i^{sption} of the HW model.

	Bermudan	vanilla swaptions	co-initial swaps	portfolio
$\alpha = 0.05$	199	-292	-8	-101
$\alpha = 0.3$	196	-299	-7	-110
$\alpha = 5$	195	-282	-6	-93
One step cov	196	-282	-6	-92

Table 3.21: Scaled total gammas of the one step and weighted covariance portfolios.

We now consider the gamma vectors which are calculated by the analogue of (3.25) for the relevant instruments. The results are displayed in figure 3.9. Again, we observe that the Bermudan's gamma vectors are very close across models. The after-hedge gamma vectors, on the other hand, seem to be quite variant. On average, Γ_i's of the one step and weighted covariance portfolios are seen to be lower in magnitude than those of the HW portfolio, especially the last gamma. This is a big improvement of the one step and weighted covariance models over the HW model.

[2]The "row sum" in this discussion corresponds to the row of the swaption matrix, NOT the row of the gamma matrix.

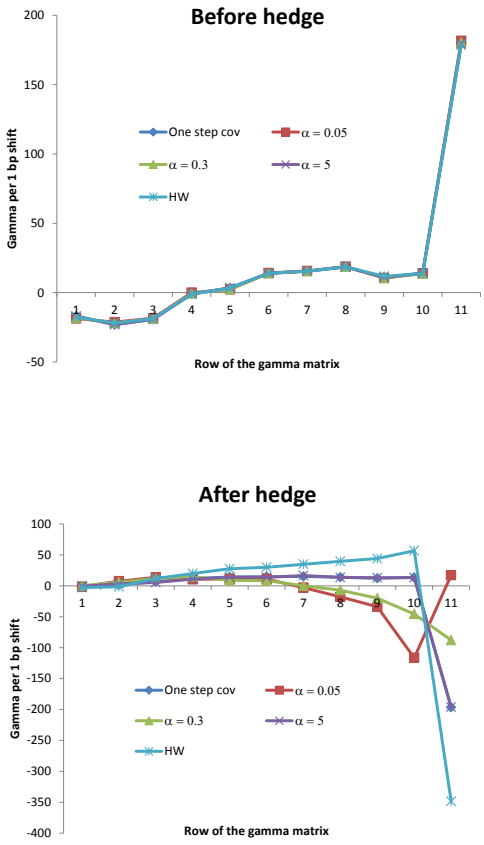


Figure 3.9: Scaled gamma vectors of the HW and the one step and weighted covariance portfolios before and after the hedge.

We again look at the gamma contribution vector from each hedging component of the portfolio in figure 3.10. As expected for all models, the magnitude of each co-ordinate of the vanilla swaptions' gamma vector dominates that of the co-initial swaps and clearly determines the gamma behaviour of the whole portfolio. Hence, we will focus on the gamma behaviour of the vanilla swaptions only. We observe that on average the contributions from the vanilla swaptions of the one step and weighted covariance portfolios seem to be much lower in magnitude than those from the co-terminal swaptions of the HW portfolio. Observe that as we move away from the reverse diagonal entries and assign more weight to other entries of the swaption matrix, the magnitude of each co-ordinate of the gamma vector can be significantly reduced.

We further observe a trend in the vanilla swaptions' gamma contributions as the parameter α of the weighted covariance model varies. When $\alpha = 0.05$, the gamma contribution is seen to be more evenly spread over all co-ordinates of the gamma vector compared with other values of α. As α increases, there are less

gamma contributions from the earlier co-ordinates ($i < n+1$) but the last gamma contribution gets bigger in magnitude (more negative). Note that the case $\alpha = 5$ is very similar to the one step covariance case (almost coincide as we see in the plots). This is because the two models have very similar vega profiles and $N_{i,j}^{\text{sption}}$'s from the vega hedge. Finally, an important point that we want to stress here is that with the one step and weighted covariance models the swaption's holder has the flexibility to control the gamma vector of the portfolio just by simply tuning the geometric weight parameter α.

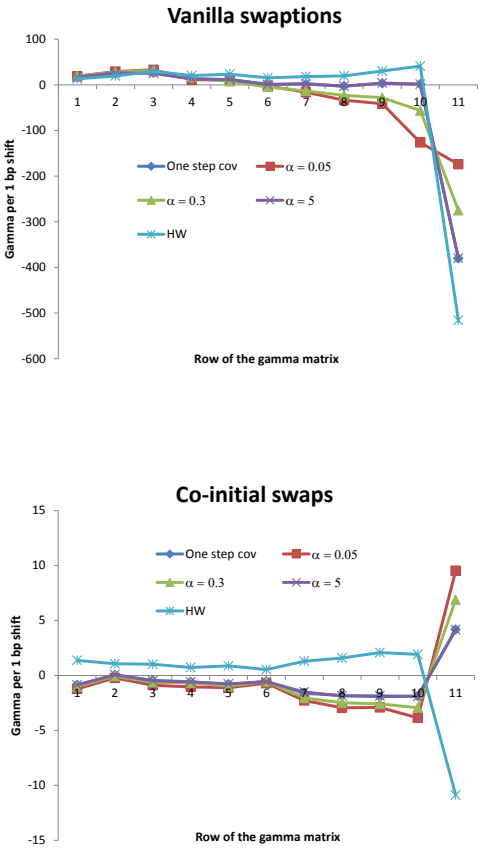


Figure 3.10: Scaled gamma contribution vectors of the vanilla swaptions and the co-initial swaps for the HW and the one step and weighted covariance portfolios.

3.6 Gamma-theta balance

In the previous section, we looked at the parallel and total gamma profiles of the hedging portfolios for different parametrizations. We recall that the total gamma of the MR portfolio is negative and almost twice as big in magnitude compared with other portfolios. Although this implies that the MR model is worse than the

others in terms of the gamma profiles, it is not the full story. Large negative gamma will only be a problem if it is not compensated by theta which is the Greek with respect to the time decay of an option. In this section, we analyze the gamma-theta balance for the hedging portfolio of the Bermudan. See Chapter 25, Volume 3 of Andersen and Piterbarg [2010] for a detailed discussion on this topic. Generally speaking, suppose we want to evaluate the change in value of the hedging portfolio as time advances forward by 1 trading day given a certain size of perturbation for all co-initial swap rates. By Taylor expansion, we have that

$$V_{t+h}^{\text{port}} \approx V_t^{\text{port}} + \frac{\partial V_t^{\text{port}}}{\partial t} h + \mathbf{\Delta}(t)\epsilon + \frac{1}{2}\epsilon^\top \mathbf{A}(t)\epsilon, \tag{3.26}$$

where h equals 1 trading day, $\epsilon^\top$ is the row vector of perturbation for all co-initial swap rates, $\mathbf{\Delta}(t)$ is the time-t delta vector and $\mathbf{A}(t)$ is the time-t gamma matrix.

Note that the term $\frac{\partial V_t^{\text{port}}}{\partial t}$ is referred to as the theta and it can be calculated numerically. Since we assume that the proportions of hedging instruments do not change over the time period $[t, t+h]$, the change in the portfolio's value purely comes from the change in the Bermudan and hedging instruments' values. As we delta hedge the Bermudan at time t, the vector $\mathbf{\Delta}(t)$ is a zero vector and hence $\mathbf{\Delta}(t)\epsilon$ will not contribute to the change in the portfolio's value. It is then clear that the theta and the gamma of the portfolio should provide a good balance so that V_{t+h}^{port} will not differ too much from V_t^{port}, i.e. we would want that $\frac{\partial V_t^{\text{port}}}{\partial t} h + \frac{1}{2}\epsilon^\top \mathbf{A}(t)\epsilon$ is as close to zero as possible (gamma-theta balance). Ideally, investors would like the sum $\frac{\partial V_t^{\text{port}}}{\partial t} h + \frac{1}{2}\epsilon^\top \mathbf{A}(t)\epsilon$ to be positive to ensure that the portfolio is not losing its value.

In practice, it is not unusual to observe that all co-initial swap rates will move up (or down) together so we will fix the perturbation to be of the same sign and size for all rates. The term $\epsilon^\top \mathbf{A}(t)\epsilon$, thus, has a close link to the total gamma as displayed below.

	Bermudan	vanilla swaptions	co-initial swaps	portfolio
MR	192	-369	-7	-184
HW	196	-286	2	-88
$\alpha = 0.05$	199	-292	-8	-101
$\alpha = 0.3$	196	-299	-7	-110
$\alpha = 5$	195	-282	-6	-93
One step cov	196	-282	-6	-92

Table 3.22: Contribution of the gamma $\epsilon^\top \mathbf{A}(t)\epsilon$ to the change in values of the portfolios as all co-initial swap rates move up (or down) by 1 bp, i.e. $\epsilon^\top = (\pm 0.0001, \ldots, \pm 0.0001)$.

We now display the theta of the portfolio including the Bermudan and its hedging instruments in the following tables. It is clear from the results that we get the scenario: long theta (positive), short gamma (negative) for the HW, one step and weighted covariance portfolios. However, for the MR portfolio we obtain the scenario: short theta and short gamma. This means that only the HW, one step and weighted covariance portfolios tend to have the gamma-theta balance. In a practical sense, the gamma-theta balance has a direct implication. A long theta-short gamma position makes money in calm markets or small ϵ in magnitude equivalently, but loses money in volatile markets or big ϵ in magnitude. A short theta-long gamma position will do the opposite. The case of a short theta-short gamma position as implied by the MR portfolio will lose money in all market scenarios.

	Bermudan	vanilla swaptions	co-initial swaps	portfolio
MR	-1806	1774	0	-32
HW	-1119	1412	0	293
$\alpha = 0.05$	-1162	1403	0	241
$\alpha = 0.3$	-1132	1447	0	315
$\alpha = 5$	-1098	1392	0	294
One step cov	-1091	1391	0	300

Table 3.23: The change in values of the portfolios as time advances by 1 trading day (theta).

We now look further into why the MR portfolio has negative theta while others have the opposite. It is seen from the tables that the theta contribution from vanilla swaptions of the MR portfolio is positive and even bigger in size than those of other portfolios. The main factor that leads to negative theta of the MR portfolio is that the Bermudan's theta following this parametrization is a lot more negative compared with others. While the Bermudan of the MR model loses 1806 after each trading day, the equivalent figures for other models is only roughly 1100. The reason for this is also clear. In order to compute the theta by finite difference, one will have to move time forward (i.e. time to expiry will decrease equivalently) and keep all other market data (implied volatilities and discount bonds) the same since theta is the partial derivative with respect to time. For example, we calculate the theta of the European swaption with today's value $\widetilde{V}_{T_0}^{i,j}$ by finite difference as $\frac{\widetilde{V}_{T_0+h}^{i,j} - \widetilde{V}_{T_0}^{i,j}}{h}$ with $\widetilde{\sigma}_{i,j}$ and $y_0^{i,j}$ both kept fixed. See also Chapter 22, Andersen and Piterbarg [2010] for different ways considered by practitioners to calculate the theta.

However, we note that computing the European swaption's value with time to maturity $T_i - h$ and implied volatility $\widetilde{\sigma}_{i,j}$ is equivalent to the one with time to expiry T_i and implied volatility $\widetilde{\sigma}_{i,j}\sqrt{\frac{T_i-h}{T_i}}$ as clearly seen from the Black formula. There-

fore, computing the theta for the Bermudan and all European swaptions by finite difference can be done equivalently by adjusting all the involved implied volatilities in the swaption matrix: $\widetilde{\sigma}_{i,j} \to \widetilde{\sigma}_{i,j}\sqrt{\frac{T_i-h}{T_i}}$ for all i, j and keep the discount bonds and time to expiry the same. This implies that the shorter the expiry is, the more the corresponding implied volatilities decrease. This adjustment of implied volatilities will not affect the correlation structure of the MR model but it will change that of the HW model as well as the one step and weighted covariance models immediately (similar to the vega analysis). Specifically, the correlation of any pair of co-terminal swap rates at their setting dates tends to decrease, e.g. this can be seen easily from the parametrization of the HW process. As a result, the Bermudan price produced by the HW, one step and weighted covariance models will decrease less than the MR model as time advances forward. This is again another interesting feature that differentiates the parametrizations by time and by expiry.

We now display the full result for the portfolios of payer Bermudan swaptions with other strikes in the following tables. We only show the results for the one step covariance case as similar conclusions are drawn for the weighted covariance model. Overall, we observe that one always gets a better gamma-theta balance for the HW and one step covariance models compared with the MR model. Note that when the strike is low (3% and 4%), we have the theta-gamma balance for the MR model but the theta is relatively small indicating that it might not be sufficient to offset the gamma risk when the market is volatile. For the large strike case (7%), a long gamma-long theta position is observed for the HW and one step covariance model indicating that the portfolios' values tend to increase regardless of market movements.

		Strike				
Model	Component	3%	4%	5%	6%	7%
MR	Bermudan	-23	207	192	118	73
	vanilla swaptions	-53	-568	-369	-168	-79
	co-initial swaps	49	14	-7	-8	-7
	portfolio	-27	-347	-184	-58	-13
HW	Bermudan	-25	211	196	117	72
	vanilla swaptions	-30	-413	-286	-135	-65
	co-initial swaps	50	21	2	-1	0
	portfolio	-5	-181	-88	-19	7
One step cov	Bermudan	-29	217	196	116	71
	vanilla swaptions	-23	-413	-282	-130	-60
	co-initial swaps	51	17	-6	-9	-7
	portfolio	-1	-179	-92	-23	4

Table 3.24: Contribution of the gamma $\epsilon^\top \mathbf{A}(t)\epsilon$ to the change in values of the portfolios of payer Bermudans with different strikes as all co-initial swap rates move up (or down) by 1 bp, i.e. $\epsilon^\top = (\pm 0.0001, \ldots, \pm 0.0001)$.

		Strike				
Model	Component	3%	4%	5%	6%	7%
MR	Bermudan	-243	-2684	-1806	-866	-445
	vanilla swaptions	250	2685	1774	846	428
	co-initial swaps	0	0	0	0	0
	portfolio	7	1	-32	-20	-17
HW	Bermudan	-83	-1301	-1119	-631	-350
	vanilla swaptions	142	1969	1412	716	373
	co-initial swaps	0	0	0	0	0
	portfolio	59	668	293	85	23
One step cov	Bermudan	-62	-1262	-1091	-621	-346
	vanilla swaptions	107	1969	1391	695	357
	co-initial swaps	0	0	0	0	0
	portfolio	45	707	300	74	11

Table 3.25: The change in values of the portfolios of payer Bermudan swaptions with different strikes as time advances by 1 trading day (theta).

As estimated from the data, it is observed that one standard deviation move per day for swap rates is roughly $\pm$ 3 bp. We display the movement of the portfolios with respect to this scenario in the table below. Given this level of movement of the

co-initial swap rates, we can see that the values of the HW and one step covariance portfolios always increase whilst that of the MR portfolio decreases with certainty.

		Strike				
Model	Component	3%	4%	5%	6%	7%
MR	Bermudan	-278	-2374	-1517	-690	-335
	vanilla swaptions	171	1833	1220	594	309
	co-initial swaps	74	20	-10	-12	-10
	portfolio	-33	-521	-307	-108	-36
HW	Bermudan	-121	-984	-825	-456	-243
	vanilla swaptions	97	1350	984	513	276
	co-initial swaps	76	31	2	-1	0
	portfolio	52	397	161	56	33
One step cov	Bermudan	-105	-936	-798	-448	-239
	vanilla swaptions	73	1350	969	500	266
	co-initial swaps	76	25	-9	-13	-11
	portfolio	44	439	162	39	16

Table 3.26: The gamma-theta balance ($\frac{\partial V_t^{\text{port}}}{\partial t}h+\frac{1}{2}\epsilon^\top \mathbf{A}(t)\epsilon$) of the portfolios of payer Bermudan swaptions with different strikes as co-initial swap rates move 3 bp after 1 trading day, i.e. $h = 1$ trading day and $\epsilon^\top = (\pm 0.0003, \ldots, \pm 0.0003)$.

3.7 Conclusions

This chapter has developed a new framework for the choice of driving process for the one-dimensional SMF model. Our approach is motivated by the problem of pricing and hedging Bermudan swaptions. We retain the computational benefit of a single-factor model but attempt to incorporate the information from a multi-factor world. It turns out that the choice of driving process x has a strong impact on the hedging behaviour of the model. In terms of the existing choices, when we construct a vega-delta neutral portfolio for a Bermudan swaption the HW model gives a much lower total gamma in magnitude compared with the MR model. The reason was found to be the fundamental difference between their imposed parametrizations by time and by expiry which leads to the difference in their vega profiles. We analyzed this issue in detail and concluded that the former outperforms the latter. The HW model, however, still lacks some flexibility in terms of the control over the parallel gamma vector of the vega-delta neutral portfolio. The main reason for this weakness was found to be the fact that the HW model only assigns the vega responses to the reverse diagonal of the swaption matrix.

In this chapter, we introduce the one step and weighted covariance models

which are different examples of parametrization by time. We observed that the new models give a very similar quality of hedge to the HW model in terms of the total gamma of vega-delta neutral portfolio. Additionally, they have an extra flexibility of the exponentially decaying weights that helps with the control over response to changes in the swaption matrix. For a certain choice of weights, the weighted covariance model spreads the vega responses over the swaption matrix and consequently reduces the magnitudes of all co-ordinates of the parallel gamma vectors of the vega-delta neutral portfolio. This is an advantage in practice. In order to support the conclusion on the gamma, we considered the gamma-theta balance of all different hedging portfolios to partly assess their potential Profit and Loss accounts. It was found that the MR portfolio tends to have a wrong gamma-theta balance which is worse than that of the HW and other portfolios associated with parametrization by time.

We believe that the driving process x plays a fundamental role in evaluating any other product and application. Furthermore, the underlying parametrization that x imposes should be one of the first criteria to consider for practitioners. It is promising that a parametrization by time process can be used in a wider context.

3.A Appendix: Estimating the market implied covariance/correlation structure

In this appendix, we show how we estimate the covariances of the log of the co-terminal forward swap rates at their setting dates from the market. The estimation is carried out in two stages using a full rank multi-factor LIBOR Market model (LMM). We first describe how to approximate the correlations of the log-LIBORs at each exercise date by a global fit to the swaption matrix. With the knowledge of these correlations, we deduce the corresponding covariances of the log-LIBORs by performing a local fit to each row of the swaption matrix. The final stage of the approximation is to use the covariances of the log-LIBORs at each exercise date to determine the target quantities, $\mathrm{Cov}(\ln y_{T_i}^{i,n+1-i}, \ln y_{T_j}^{j,n+1-j})$ for $i = 1, \ldots, n-1$ and $j > i$. This can be done by using the relationship between the swap rates and the LIBORs.

3.A.1 Approximating the terminal correlations, a global fit approach

We first introduce the n-factor LMM under the terminal measure $\mathbb{S}^{n+1}$. Suppose we are given n deterministic instantaneous volatility functions $\sigma_i(t), t \leq T_i$ for each $i = 1, \ldots, n$. We further introduce the instantaneous correlation $\rho_{ij} \in [-1, 1]$ for

each pair of Brownian factors W^i and W^j, i.e. $\mathrm{d}W_t^i \mathrm{d}W_t^j = \rho_{ij}\mathrm{d}t$. Under the terminal measure $\mathbb{S}^{n+1}$, the n-factor LMM reads

1. $i = n, t \leq T_n$:

$$\mathrm{d}L_t^n = \sigma_n(t) L_t^n \mathrm{d}W_t^n,$$

2. $i < n, t \leq T_i$:

$$\mathrm{d}L_t^i = -\left(\sum_{j=i+1}^{n} \frac{\alpha_j L_t^j}{1+\alpha_j L_t^j} \sigma_i(t)\sigma_j(t)\rho_{ij} \right) L_t^i \mathrm{d}t + \sigma_i(t) L_t^i \mathrm{d}W_t^i. \tag{3.27}$$

The formal solution to the SDE (3.27) is

$$L_t^i = L_0^i \times \exp\left(\int_0^t \sum_{j=i+1}^{n} \frac{\alpha_j L_s^j}{1+\alpha_j L_s^j} \sigma_i(s)\sigma_j(s)\rho_{ij}\mathrm{d}s - \int_0^t \frac{\sigma_i^2(s)}{2}\mathrm{d}s + \int_0^t \sigma_i(t)\mathrm{d}W_t^i \right).$$

As the drift terms are stochastic, one usually seeks for a fast numerical scheme to approximate the solution. In the current literature, by "freezing the drift" at time zero we obtain a very crude drift approximation which could encourage quite a significant arbitrage

$$L_t^i \approx L_0^i \times \exp\left(\sum_{j=i+1}^{n} \frac{\alpha_j L_0^j}{1+\alpha_j L_0^j} \int_0^t \sigma_i(s)\sigma_j(s)\rho_{ij}\mathrm{d}s - \int_0^t \frac{\sigma_i^2(s)}{2}\mathrm{d}s + \int_0^t \sigma_i(t)\mathrm{d}W_t^i \right). \tag{3.28}$$

However, it is particularly useful when calibrating the model to the "terminal correlation" because it allows for an analytically closed form formula. Here, by "terminal correlation" we mean the correlation between the log-LIBORs at the same setting date, e.g. $\mathrm{Corr}(\ln L_{T_i}^k, \ln L_{T_i}^l)$ for $k, l \geq i$. Similar versions for other setting dates hold as well, see for example Brigo and Mercurio [2001] or Rebonato [2004]. By (3.28), we get

$$\mathrm{Corr}(\ln L_{T_i}^k, \ln L_{T_i}^l) \approx \frac{\int_0^{T_i} \sigma_k(t)\sigma_l(t)\rho_{kl}\mathrm{d}t}{\sqrt{\int_0^{T_i} \sigma_k^2(t)\mathrm{d}t}\sqrt{\int_0^{T_i} \sigma_l^2(t)\mathrm{d}t}}, \qquad k, l \geq i. \tag{3.29}$$

The above quantities can only be approximated given the parameters for the instantaneous volatilities and the instantaneous correlations.

Instantaneous correlation: we use the following simple and financially appealing form

$$\rho_{ij} = \rho_{ij}(\beta) = \exp(-\beta|T_i - T_j|),$$

where $\beta > 0$. For more details of this choice, readers are referred to Rebonato

[2002, 2004] and Brigo and Mercurio [2001].

Instantaneous volatility: we choose the "humped shape" function, originally proposed by Rebonato [2002, 2004]

$$\sigma_i(t) = \sigma_i(t; a, b, c, d) = [a + b(T_i - t)]e^{-c(T_i - t)} + d,$$

where $a, b, c, d \in \mathbb{R}$ are the four parameters to be chosen appropriately.

A global fit: we now aim to recover all the parameters of the n-factor LMM model in order to approximate the terminal correlations in (3.29). This can be done by performing a global fit to the swaption matrix which contains implied volatilities of different expiries and tenors.

Note that each swap rate can be written as

$$\begin{aligned} y_t^{i,j} &= \sum_{k=i}^{i+j-1} w_k^{i,j}(t) L_t^k, \\ w_k^{i,j}(t) &= \frac{\alpha_k D_{tT_{k+1}}}{\sum_{l=i}^{i+j-1} \alpha_l D_{tT_{l+1}}}, \end{aligned}$$

By Itô's lemma, under the terminal measure we have that

$$\begin{aligned} \mathrm{d}y_t^{i,j} &= \ldots \mathrm{d}t + \left(\sum_{k=i}^{i+j-1} \zeta_k^{i,j}(t) \sigma_k(t) \mathrm{d}W_t^k \right) y_t^{i,j}, \qquad (3.30) \\ \zeta_k^{i,j}(t) &= \frac{\widetilde{w}_k^{i,j}(t) L_t^k}{y_t^{i,j}}, \\ \widetilde{w}_k^{i,j}(t) &= w_k^{i,j}(t) + \sum_{l=i, l\neq k}^{i+j-1} \frac{\partial w_l^{i,j}(t)}{\partial L_t^k} L_t^l. \end{aligned}$$

Hence, if the corresponding swaption is to be valued using the Black formula with implied volatility $\widetilde{\sigma}_{i,j}$, we would want to get

$$\widetilde{\sigma}_{i,j}^2 T_i = \sum_{k=i}^{i+j-1} \sum_{l=i}^{i+j-1} \int_0^{T_i} \zeta_k^{i,j}(t) \zeta_l^{i,j}(t) \sigma_k(t) \sigma_l(t) \rho_{kl} \mathrm{d}t,$$

by ignoring the drift in (3.30) and assuming that the terms $\zeta_k^{i,j}(t)$ are all deterministic. Empirically, as shown in Rebonato [2002], Rebonato [2004] and Brigo and Mercurio [2001], one can actually obtain that each $\zeta_k^{i,j}(t)$ is approximately equal to its value today, i.e. $\zeta_k^{i,j}(0)$. The above equation now becomes of a much simpler form

$$\widetilde{\sigma}_{i,j}^2 T_i \approx \sum_{k=i}^{i+j-1} \sum_{l=i}^{i+j-1} \zeta_k^{i,j}(0) \zeta_l^{i,j}(0) \int_0^{T_i} \sigma_k(t) \sigma_l(t) \rho_{kl} \mathrm{d}t. \qquad (3.31)$$

This allows us to a carry a fast yet accurate enough approximation scheme for both the global and the local fits as we shall see later.

We can now use a least squares fit method to do a global fit to the swaption matrix. For a particular choice of parameters $\{\beta, a, b, c, d\}$, we define the model volatilities $\{\sigma_{i,j}(\beta; a, b, c, d)\}$ to satisfy

$$\sigma_{i,j}^2(\beta; a, b, c, d)T_i := \sum_{k=i}^{i+j-1} \sum_{l=i}^{i+j-1} \zeta_k^{i,j}(0)\zeta_l^{i,j}(0) \int_0^{T_i} \sigma_k(t; a, b, c, d)\sigma_l(t; a, b, c, d)\rho_{kl}(\beta)\mathrm{d}t.$$

One then defines

$$\chi^2 := \sum_{i=1}^{n} \sum_{j=1}^{n+1-i} \left[\sigma_{i,j}(\beta; a, b, c, d) - \widetilde{\sigma}_{i,j}\right]^2,$$

and looks for the optimal set of parameters $\{\beta, a, b, c, d\}$ that minimizes χ^2. At the end of this stage, we will have recovered all the parameters of the n-factor LMM and hence the terminal correlations in (3.29) can be estimated.

Remark 7 *Note that in this step, we do not take the value* $\int_0^{T_i} \sigma_k(t)\sigma_l(t)\rho_{kl}(\beta)\mathrm{d}t$ *as an approximation for Cov*$(\ln L_{T_i}^k, \ln L_{T_i}^l)$ *(the global approach does not reflect enough accuracy). The local step presented next will give a better approximation for these covariances.*

We assume the correlations between the log-LIBORs obtained from the global fit are not affected by any changes in the market data. This is because the changes in these correlations are recorded to be small and do not have a big impact on the approximations from the local fit in Appendix 3.A.2. In our check, we find that the effect is numerically insignificant and this suggests that historical data can be used for the global fit. Hence, we keep the parameters of the instantaneous volatility and the instantaneous correlation functions the same at all time.

3.A.2 Approximating the covariances, a local fit approach

Recall that from the global fit we can approximate the correlations $\mathrm{Corr}(\ln L_{T_i}^k, \ln L_{T_i}^l)$ for $k, l \geq i$ at any exercise date T_i. In order to deduce the corresponding covariances of the log-LIBORs at T_i, we use the implied volatilities on the i$^{\text{th}}$ row of the swaption matrix. We employ the approximation in (3.31) but use $\mathrm{Cov}(\ln L_{T_i}^k, \ln L_{T_i}^l)$ instead of $\int_0^{T_i} \sigma_k(t)\sigma_l(t)\rho_{kl}\mathrm{d}t$, i.e.

$$\widetilde{\sigma}_{i,j}^2 T_i \approx \sum_{k=i}^{i+j-1} \sum_{l=i}^{i+j-1} \zeta_k^{i,j}(0)\zeta_l^{i,j}(0)\mathrm{Cov}(\ln L_{T_i}^k, \ln L_{T_i}^l). \tag{3.32}$$

The following approximation steps are described to solve for $\mathrm{Var}(\ln L_{T_i}^j)$ where j runs from i to n. Since we keep all the correlations fixed, solving for the variances

automatically implies the covariances. Note that we ignore the effect of changing the measure as it is small and irrelevant for the discussion.

- **Step 1**: we start from $\widetilde{\sigma}_{i,1}$ of the i$^{\text{th}}$ caplet. The Black formula implies that $\widetilde{\sigma}_{i,1}^2 T_i \approx \mathbf{Var}(\ln \mathbf{L}_{\mathbf{T_i}}^{\mathbf{i}})$.

- **Step 2**: next, we consider $\widetilde{\sigma}_{i,2}$. Also, by the Black formula and the approximation in (3.32)

$$\begin{aligned}
\widetilde{\sigma}_{i,2}^2 T_i &\approx (\zeta_i^{i,2}(0))^2 \mathrm{Var}(\ln L_{T_i}^i) \\
&\quad +(\zeta_{i+1}^{i,2}(0))^2 \mathbf{Var}(\ln \mathbf{L}_{\mathbf{T_i}}^{\mathbf{i+1}}) + 2\zeta_i^{i,2}(0)\zeta_{i+1}^{i,2}(0)\mathrm{Cov}(\ln L_{T_i}^i, \ln L_{T_i}^{i+1}) \\
&= (\zeta_i^{i,2}(0))^2 \mathrm{Var}(\ln L_{T_i}^i) + (\zeta_{i+1}^{i,2}(0))^2 \mathbf{Var}(\ln \mathbf{L}_{\mathbf{T_i}}^{\mathbf{i+1}}) \qquad (3.33) \\
&\quad +2\zeta_i^{i,2}(0)\zeta_{i+1}^{i,2}(0)\mathrm{Corr}(\ln L_{T_i}^i, \ln L_{T_i}^{i+1}) \sqrt{\mathrm{Var}(\ln L_{T_i}^i)} \sqrt{\mathbf{Var}(\ln \mathbf{L}_{\mathbf{T_i}}^{\mathbf{i+1}})}
\end{aligned}$$

  It is straightforward to solve this equation for the unknown $\mathbf{Var}(\ln \mathbf{L}_{\mathbf{T_i}}^{\mathbf{i+1}})$. Hence, $\mathrm{Cov}(\ln L_{T_i}^i, \ln L_{T_i}^{i+1})$ can be recovered.

- **Step j+1-i**: notice that each time we move from $\widetilde{\sigma}_{i,j-i}$ to $\widetilde{\sigma}_{i,j+1-i}$, there is one more unknown to solve, i.e. $\mathbf{Var}(\ln \mathbf{L}_{\mathbf{T_i}}^{\mathbf{j}})$. With the knowledge of the terminal correlation, we can recover $\mathrm{Cov}(\ln L_{T_i}^k, \ln L_{T_i}^j)$ for $k = i, \ldots, j-1$. This is clear by the following relation

$$\begin{aligned}
&\widetilde{\sigma}_{i,j+1-i}^2 T_i \\
\approx\ & \sum_{k=i}^{j} \sum_{l=i}^{j} \zeta_k^{i,j+1-i}(0)\zeta_l^{i,j+1-i}(0)\mathrm{Cov}(\ln L_{T_i}^k, \ln L_{T_i}^l) \\
=\ & \cdots + (\zeta_j^{i,j+1-i}(0))^2 \mathbf{Var}(\ln \mathbf{L}_{\mathbf{T_i}}^{\mathbf{j}}) \qquad (3.34) \\
& +2\sum_{k=i}^{j-1} \zeta_k^{i,j+1-i}(0)\zeta_j^{i,j+1-i}(0)\mathrm{Cov}(\ln L_{T_i}^k, \ln L_{T_i}^j) \\
=\ & \cdots + (\zeta_j^{i,j+1-i}(0))^2 \mathbf{Var}(\ln \mathbf{L}_{\mathbf{T_i}}^{\mathbf{j}}) \qquad (3.35) \\
& +2\sum_{k=i}^{j-1} \zeta_k^{i,j+1-i}(0)\zeta_j^{i,j+1-i}(0)\mathrm{Corr}(\ln L_{T_i}^k, \ln L_{T_i}^j) \sqrt{\mathrm{Var}(\ln L_{T_i}^k)} \sqrt{\mathbf{Var}(\ln \mathbf{L}_{\mathbf{T_i}}^{\mathbf{j}})}
\end{aligned}$$

  where the $\ldots$ terms and $\mathrm{Var}(\ln L_{T_i}^k)$, for $k = i, \ldots, j-1$ are known from the previous steps.

- **Step n+1-i**: At the end of this step, we will have recovered the variances and covariances of all the alive log-LIBORs at T_i.

For $j > i$, by using the approximation $\zeta_k^{i,j}(t) \approx \zeta_k^{i,j}(0)$ in (3.30) we have that

$$\mathrm{d}\ln y_t^{i,j} \approx \ldots \mathrm{d}t + \left(\sum_{k=i}^{i+j-1} \zeta_k^{i,j}(0)\sigma_k(t)\mathrm{d}W_t^k \right). \tag{3.36}$$

Hence, our target market quantity can be written as

$$\mathrm{Cov}(\ln y_{T_i}^{i,n+1-i}, \ln y_{T_j}^{j,n+1-j}) \approx \mathrm{Cov}(\ln y_{T_i}^{i,n+1-i}, \ln y_{T_i}^{j,n+1-j}), \qquad j > i,$$

which follows from the independence of increment in (3.36). Again, since we use $\mathrm{Cov}(\ln L_{T_i}^k, \ln L_{T_i}^l)$ instead of $\int_0^{T_i} \sigma_k(t)\sigma_l(t)\rho_{kl}\mathrm{d}t$, it follows from (3.36) that

$$\mathrm{Cov}(\ln y_{T_i}^{i,n+1-i}, \ln y_{T_i}^{j,n+1-j}) \approx \sum_{k=i}^{n}\sum_{l=j}^{n} \zeta_k^{i,n+1-i}(0)\zeta_l^{j,n+1-j}(0)\mathrm{Cov}(\ln L_{T_i}^k, \ln L_{T_i}^l). \tag{3.37}$$

When setting the various models, we will denote the covariances of the log of the swap rates $\mathrm{Cov}(\ln y_{T_i}^{i,n+1-i}, \ln y_{T_j}^{j,n+1-j})$ estimated from the market by this two step procedure as $\mathrm{Cov}^{\mathrm{ma}}(\ln y_{T_i}^{i,n+1-i}, \ln y_{T_j}^{j,n+1-j})$ for $j = i+1, \ldots, n$.

3.B Appendix: Explanations for the deltas of the Bermudan swaption

In this appendix, we give a brief and heuristic explanation for the behaviour of the Bermudan's deltas in the main chapter.

3.B.1 Delta calculation

Recall from the main chapter that the deltas of a Bermudan swaption are defined to be

$$\hat{\Delta}_i := \frac{\mathrm{d}\hat{V}_{T_0}}{\mathrm{d}y^{0,i}},$$

for $i = 1, \ldots, n+1$. This gives a vector $\hat{\mathbf{\Delta}}$ and each element $\hat{\Delta}_i$ is itself a function of the co-initial swap rates. We describe below how we carry out the finite difference scheme to calculate the deltas.

Bumping the co-initial swap rates: consider the initial discount bonds with values D_{0T_i} for $i = 1, \ldots, n+1$. The co-initial swap rates and discount bonds are linked via the relation

$$y^{0,i} = \frac{1 - D_{0T_i}}{\sum_{j=0}^{i-1} \alpha_j D_{0T_j+1}}, \qquad i = 1, \ldots, n+1. \tag{3.38}$$

We can also calculate each discount bond D_{0T_i} from the co-initial swap rate $y^{0,i}$ and

the discount bonds that mature previously: $D_{0T_1}, \ldots, D_{0T_{i-1}}$

$$D_{0T_i} = \frac{1 - y^{0,i} \sum_{j=0}^{i-2} \alpha_j D_{0T_{j+1}}}{1 + \alpha_{i-1} y^{0,i}}. \tag{3.39}$$

This is because if we calculate the right hand side of (3.39) by using (3.38), we will have that

$$\begin{aligned}
\frac{1 - y^{0,i} \sum_{j=0}^{i-2} \alpha_j D_{0T_{j+1}}}{1 + \alpha_{i-1} y^{0,i}} &= \frac{1 - \frac{1 - D_{0T_i}}{\sum_{j=0}^{i-1} \alpha_j D_{0T_j+1}} \sum_{j=0}^{i-2} \alpha_j D_{0T_{j+1}}}{1 + \alpha_{i-1} y^{0,i}} \\
&= \frac{1 - \frac{1 - D_{0T_i}}{\sum_{j=0}^{i-1} \alpha_j D_{0T_j+1}} \left(\sum_{j=0}^{i-1} \alpha_j D_{0T_{j+1}} - \alpha_{i-1} D_{0T_i} \right)}{1 + \alpha_{i-1} y^{0,i}} \\
&= \frac{1 - (1 - D_{0T_i}) + D_{0T_i} \frac{\alpha_{i-1} - \alpha_{i-1} D_{0T_i}}{\sum_{j=0}^{i-1} \alpha_j D_{0T_j+1}}}{1 + \alpha_{i-1} y^{0,i}} \\
&= \frac{D_{0T_i} + D_{0T_i} \frac{\alpha_{i-1} - \alpha_{i-1} D_{0T_i}}{\sum_{j=0}^{i-1} \alpha_j D_{0T_j+1}}}{1 + \alpha_{i-1} y^{0,i}} \\
&= D_{0T_i} \frac{1 + \alpha_{i-1} \frac{1 - D_{0T_i}}{\sum_{j=0}^{i-1} \alpha_j D_{0T_j+1}}}{1 + \alpha_{i-1} y^{0,i}} \\
&= D_{0T_i} \frac{1 + \alpha_{i-1} y^{0,i}}{1 + \alpha_{i-1} y^{0,i}} \\
&= D_{0T_i}.
\end{aligned}$$

It is then straightforward to shift $y^{0,i}$ and keep $y^{0,j}, j \neq i$ constant for each i.

3.B.2 Main effects of bumping the co-initial swap rates on the inputs and the implication for the deltas

We now explain how bumping the co-initial swap rates can affect the inputs and the Bermudan price. We display the LIBORs and the co-terminal swap rates in tables 3.27 and 3.28 respectively. The second columns of both tables display the rates before the bump of each co-initial swap rate. Other columns show their updated values after the bump.

We bump each of $y^{0,i}$ and keep the rest unchanged to calculate the deltas as described in the previous section. When $i \leq n$, it follows directly from (3.39) that only $\{D_{0T_j}\}_{j=i,\ldots,n+1}$ alter. D_{0T_i} will decrease as $y^{0,i}$ increases but $D_{0T_{i+1}}$ will increase to keep $y^{0,i+1}$ unchanged (see (3.39)). This will have a direct effect on the two consecutive LIBORs

$$L_0^{i-1} = \frac{D_{0T_{i-1}} - D_{0T_i}}{\alpha_{i-1} D_{0T_i}} \quad \text{increases,} \qquad L_0^i = \frac{D_{0T_i} - D_{0T_{i+1}}}{\alpha_i D_{0T_{i+1}}} \quad \text{decreases.}$$

The later discount bonds are also affected but it is observed numerically that the effects on the LIBORs for the later periods are very small (table 3.27).

As we can expect, a change of the LIBORs will directly lead to a change of the involved co-terminal swap rates. Since one can view a co-terminal swap rate approximately as a linear combination of the corresponding spanning LIBORs, bumping $y^{0,i}$ for $i \leq n$ effectively has only one main effect on $y_0^{i,n+1-i}$. Because L_0^i decreases and the later LIBORs are almost the same as before the bump, $y_0^{i,n+1-i}$ will decrease (see table 3.28). It is clear that the co-terminal swap rates with expiries after T_i will almost stay the same as the effects on their spanning LIBORs are very small. The co-terminal swap rates with expiries before T_i are also much less affected. This is because the increase and decrease in the corresponding spanning LIBORs L_0^{i-1} and L_0^i tend to cancel each other out and hence leave these co-terminal swap rates which are the average sums almost unchanged. Note that the decrease in $y_0^{i,n+1-i}$ is seen to be relatively small and hence one can expect that it leads to a negative delta of quite small magnitude (figure 3.4).

The last delta is quite different from the others. Bumping $y^{0,n+1}$ will only decrease the terminal discount bond $D_{0T_{n+1}}$. This causes the last LIBOR and all the co-terminal swap rates to increase because

$$y_0^{i,n+1-i} = \frac{D_{0T_i} - D_{0T_{n+1}}}{\sum_{j=i}^{n} \alpha_j D_{0T_j+1}}.$$

This can cause the overall effect on the Bermudan to be quite large. The last delta $\hat{\Delta}_{n+1}$, thus, is positive and could have a very large magnitude compared with the other deltas. This explains the large jump of the last delta in figure 3.4.

Bump	None	$y^{0,1}$	$y^{0,2}$	$y^{0,3}$	$y^{0,4}$	$y^{0,5}$	$y^{0,6}$	$y^{0,7}$	$y^{0,8}$	$y^{0,9}$	$y^{0,10}$	$y^{0,11}$
L_0^0	4.54	**4.55**	4.54	4.54	4.54	4.54	4.54	4.54	4.54	4.54	4.54	4.54
L_0^1	4.54	**4.53**	**4.56**	4.54	4.54	4.54	4.54	4.54	4.54	4.54	4.54	4.54
L_0^2	4.55	4.55	**4.53**	**4.58**	4.55	4.55	4.55	4.55	4.55	4.55	4.55	4.55
L_0^3	4.58	4.58	4.58	**4.55**	**4.63**	4.58	4.58	4.58	4.58	4.58	4.58	4.58
L_0^4	4.65	4.65	4.65	4.65	**4.61**	**4.71**	4.65	4.65	4.65	4.65	4.65	4.65
L_0^5	4.71	4.71	4.71	4.71	4.71	**4.65**	**4.78**	4.71	4.71	4.71	4.71	4.71
L_0^6	4.80	4.80	4.80	4.80	4.80	4.80	**4.73**	**4.88**	4.80	4.80	4.80	4.80
L_0^7	4.89	4.89	4.89	4.89	4.89	4.89	4.89	**4.80**	**4.98**	4.89	4.89	4.89
L_0^8	4.99	4.99	4.99	4.99	4.99	4.99	4.99	4.99	**4.90**	**5.10**	4.99	4.99
L_0^9	5.06	5.06	5.06	5.06	5.06	5.06	5.06	5.06	5.06	**4.95**	**5.19**	5.06
L_0^{10}	5.13	5.13	5.13	5.13	5.13	5.13	5.13	5.13	5.13	5.13	**5.00**	**5.27**

Table 3.27: Effects of bumping the co-initial swap rates by 1 bp on the LIBORs (in percentage). The bold figures represent the main effects.

Bump	None	$y^{0,1}$	$y^{0,2}$	$y^{0,3}$	$y^{0,4}$	$y^{0,5}$	$y^{0,6}$	$y^{0,7}$	$y^{0,8}$	$y^{0,9}$	$y^{0,10}$	$y^{0,11}$
$y_0^{1,10}$	4.76	**4.76**	4.76	4.76	4.76	4.76	4.76	4.76	4.76	4.76	4.76	**4.77**
$y_0^{2,9}$	4.79	4.79	**4.79**	4.79	4.79	4.79	4.79	4.79	4.79	4.79	4.79	**4.81**
$y_0^{3,8}$	4.83	4.83	4.83	**4.83**	4.83	4.83	4.83	4.83	4.83	4.83	4.83	**4.85**
$y_0^{4,7}$	4.87	4.87	4.87	4.87	**4.87**	4.87	4.87	4.87	4.87	4.87	4.87	**4.89**
$y_0^{5,6}$	4.92	4.92	4.92	4.92	4.92	**4.91**	4.92	4.92	4.92	4.92	4.92	**4.94**
$y_0^{6,5}$	4.97	4.97	4.97	4.97	4.97	4.97	**4.95**	4.97	4.97	4.97	4.97	**4.99**
$y_0^{7,4}$	5.01	5.01	5.01	5.01	5.01	5.01	5.01	**4.99**	5.01	5.01	5.01	**5.05**
$y_0^{8,3}$	5.06	5.06	5.06	5.06	5.06	5.06	5.06	5.06	**5.02**	5.06	5.06	**5.10**
$y_0^{9,2}$	5.09	5.09	5.09	5.09	5.09	5.09	5.09	5.09	5.09	**5.03**	5.09	**5.16**
$y_0^{10,1}$	5.13	5.13	5.13	5.13	5.13	5.13	5.13	5.13	5.13	5.13	**5.00**	**5.27**

Table 3.28: Effects of bumping the co-initial swap rates by 1 bp on the co-terminal swap rates (in percentage). The bold figures represent the main effects.

Part II

Stochastic volatility modelling

Chapter 4

An overview of smile modelling

Part II of the thesis focuses on the problem of modelling volatility smile. This chapter gives a brief overview of the problem as well as some current issues in the literature.

4.1 Black/Normal models and general market consensus

The Black-Scholes (in Black and Scholes [1973]) or Black (in Black [1976]) models are amongst the simplest models to describe the dynamics of an asset in practice. We recall from Chapter 2 that under the Black model, the underlying asset has a Log-Normal distribution that follows the driftless SDE under some martingale measure

$$\mathrm{d}F_t = \sigma F_t \mathrm{d}W_t, \qquad t \in [0, T],$$

where W is a one-dimensional Brownian motion. The constant volatility parameter $\sigma > 0$ is the main input for the model to arrive at the famous Black formula. Similar to the application of the Black model to the evaluation of European swaptions and caplets as discussed in Chapter 2, the numeraire rebased today's price of a European call option written on F with strike K and maturity T is given in closed form

$$\begin{aligned} \mathbb{C}_0(F_0, K, T) &= \mathbb{E}[(F_T - K)^+] \\ &= F_0\Phi(d_1) - K\Phi(d_2), \end{aligned}$$

where $\Phi(\cdot)$ is the standard Normal CDF and

$$\begin{aligned} d_1 &= \frac{\ln(\frac{F_0}{K}) + \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}}, \\ d_2 &= d_1 - \sigma\sqrt{T}. \end{aligned}$$

As noted in Chapter 2, conversely the market price of the above European call option is usually quoted via the Black implied volatility $\widetilde{\sigma}$ for each strike K. That is the value for the volatility parameter that we plug into the above Black formula to reproduce the market price. Hence, if we set $\sigma = \widetilde{\sigma}$ in the Black model for each strike K, the model price computed by the Black formula and the market price will coincide. One also has a choice to use a deterministic function of time $\sigma(\cdot)$ instead of the constant σ in the SDE for the underlying, but has to ensure that $\int_0^T \sigma^2(t)\mathrm{d}t = \widetilde{\sigma}^2 T$ for the perfect recovery of the market price.

In the previous chapters, we assumed that the Black implied volatility stays constant for all strikes. In fact, this is not the case in practice. After the market crash in 1987, it became more pronounced to practitioners that the log of the underlying asset at time T is far from Normal and a constant volatility parameter σ or even a deterministic function of time $\sigma(\cdot)$ are not possible. What usually features in the market is that the market Black implied volatilities that we invert from market option prices display a variety of shapes as a function of strike. The most common is the U-shape or smile (higher volatilities for low and high strikes). In interest rate markets, skew is a more dominant feature, i.e. Black implied volatility tends to be monotonically decreasing in strike and displays less convexity for large strike. Despite the fall of the Black model, Black implied volatility is still one of the market standards to quote vanilla prices in practice. This has become a common metric as to how traders think and relate prices in a trading environment.

Another way of quoting vanilla prices is via the Normal implied volatilities. The underlying model is the Normal (or Bachelier) model where we assume the dynamics of the underlying asset is given by

$$\mathrm{d}F_t = \sigma \mathrm{d}W_t.$$

The constant volatility parameter $\sigma > 0$ (or a deterministic function of time $\sigma(\cdot)$) in this model is the Normal rather than Log-Normal volatility as in the previous Black model. One can immediately obtain a closed-form expression for vanilla call and put prices in a similar fashion to the Black formula, and such expression is the so-called Bachelier formula with the analogous Normal implied volatility. In certain markets, traders tend to think in terms of the Normal implied volatilities since the Normal model sometimes captures the dynamics of the underlying asset better than the Black model. A clear drawback of this model is that F becomes negative in finite time with positive probability which is not desired. Henceforth, we note that the implied volatility that we refer to is always the Black implied volatility unless otherwise specified. We return to the Normal implied volatilities in Chapter 7 where we discuss the hedging of European options with stochastic volatility.

Certainly, the Black and Normal models are both not rich enough to capture

the market implied distribution of the underlying. In terms of the implied volatility smile, even when we equip the Black model with a time-dependent volatility function rather than just a constant, we still end up with a flat implied volatility curve at time T. The reason is that the model implied distribution of the underlying under the martingale measure is always Log-Normal which is not consistent with the market. See also Chapter 2 for a similar discussion on market implied distributions. This means that one cannot use the same model to price vanilla European options for all strikes. This leads us to a further study of richer classes of volatility models to account for the smile/skew effect.

4.2 Local volatility models

Local volatility models have been mentioned and studied in a large body of works to achieve the objectives of recovering market smiles. Here we name a few early works on local volatility models that are greatly acknowledged, e.g. Duprie [1994], Derman and Kani [1994], and Derman and Kani [1998]. See also Carr [2000] for a brief survey on the related subject. The class of local volatility models is essentially a natural extension of the Black model where one writes the dynamics of underlying asset under some martingale measure as

$$\mathrm{d}F_t = \sigma\varphi(F_t)\mathrm{d}W_t.$$

Here, we still assume the simplest case of constant volatility $\sigma > 0$ and $\varphi(\cdot)$ is a suitable deterministic function of the underlying asset that satisfies regularity conditions (see page 278, Volume 1, Andersen and Piterbarg [2010]). The introduction of function $\varphi(\cdot)$ has a direct impact on implied volatilities. If $\varphi(x)/x$ is a monotonically decreasing function in x, the smile will be a downward-sloping skew. If $\varphi(x)/x$ is non-monotonic, we will be able to obtain a true U-shaped smile. See for example Andersen and Andreasen [2002] or page 279, Volume 1, Andersen and Piterbarg [2010] for implications for smile shapes of different choices of the function $\varphi(\cdot)$. Two popular examples of local volatility models that have been used extensively in practice to model skews are described below.

CEV Model: The Constant Elasticity of Variance model, or CEV, has appeared in a number of works, e.g. Schroder [1989], Cox [1996]. In the CEV model, the local volatility function $\varphi(x) = x^\beta$ is of the power law type with the exponent β typically ranging between $[0, 1]$

$$\mathrm{d}F_t = \sigma F_t^\beta \mathrm{d}W_t.$$

When $\beta = 0$ or 1, the CEV model collapses to the Bachelier or Black models respectively. The flexibility of β was introduced to better capture the skew, i.e. more negative slope is observed as β is reduced from 1 to 0 (we will discuss this in more detail in the next chapter). Of course, this SDE formulation will never admit negative solutions. Otherwise, F_t^β will be a complex number. Depending on the value of β, we then have that F_t either absorbs or reflects at zero and its transition density is known in closed form. Practitioners, on the other hand, tend to choose the absorbing or reflecting behaviours based on the markets rather than the choice of β, e.g. absorption at zero for interest rate but reflection at zero for equity. See also Brigo and Mercurio [2001], or Volume I, Andersen and Piterbarg [2010] for option pricing in the CEV model.

Displaced Diffusion model: Another relatively simple way to account for volatility skew is to use the Displaced Diffusion (DD) model which was first introduced in Rubinstein [1983]. In this case, $\varphi(x) = x + \theta$ where θ is a positive constant

$$\mathrm{d}F_t = \sigma(F_t + \theta)\mathrm{d}W_t.$$

When $\theta = 0$, the DD model collapses to the Black model and becomes very similar to the Bachelier model as $\theta \to \infty$. Under some certain model parametrization, Marris [1999] shows that the DD and CEV models are in fact very similar qualitatively. With this SDE formulation, one again can price European calls and puts in closed form, e.g. for call option we have the following DD-Black formula

$$\begin{aligned}\mathbb{C}_0(F_0, K, T) &= \mathbb{E}[(F_T - K)^+] \\ &= (F_0 + \theta)\Phi(d_1) - (K + \theta)\Phi(d_2),\end{aligned}$$

where

$$\begin{aligned}d_1 &= \frac{\ln\left(\frac{F_0+\theta}{K+\theta}\right) + \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}}, \\ d_2 &= d_1 - \sigma\sqrt{T}.\end{aligned}$$

A clear drawback of this model is that it admits negative values for the underlying asset with positive probability. We will discuss this issue in more detail in Chapters 5 and 6. For a more complete reference to the CEV and DD models, readers are referred to several books such as Musiela and Rutkowski [2004], Brigo and Mercurio [2001], Andersen and Piterbarg [2010], Rebonato [2004], etc.

4.3 Stochastic volatility models

We mentioned that the class of local volatility models is capable of producing skews/smiles exhibited in various markets. The fact that the model has only one factor enables a fairly simple and straightforward implementation in practice. However, this simplicity comes with a cost. Hagan et al. [2002] was the first to question whether local volatility models match the qualitative behaviour of the markets. In Hagan et al. [2002], the authors show that no matter how well we fit to market smiles, local volatility models are still not good enough to capture the full dynamics of the underlying. Specifically, it was pointed out by the authors that local volatility models predict smiles to move with the underlying in an "opposite" direction to the market. However, this is based on a rather "naive" hedging concept which is not always practically useful. In Chapter 7, we will discuss different hedging methodologies in detail.

It is interesting to note that stochastic volatility models appeared long before local volatility models were criticized. They were, in fact, studied at the time when analytical solutions and computer power for simulation were not always available for practical use. In a general stochastic volatility model, volatility is a stochastic process itself with a separate Brownian motion Z to describe the randomness in addition to the SDE for the underlying asset. A general SDE for the volatility process is usually of the form

$$\mathrm{d}\sigma_t = g(\sigma_t)\mathrm{d}t + h(\sigma_t)\mathrm{d}Z_t,$$

where $g(\cdot)$ and $h(\cdot)$ describe the drift and diffusion terms respectively. The literature for stochastic volatility models is vast. One of the first contributions in this area is the model proposed in Hull and White [1987] where the underlying asset has the usual Black-Scholes (Log-Normal) SDE and the volatility dynamics follows the driftless Log-Normal SDE

$$\mathrm{d}\sigma_t = \nu\sigma_t\mathrm{d}Z_t, \qquad \nu > 0.$$

Here, volatility is assumed to be independent of the underlying, i.e. two Brownian motions W and Z are uncorrelated. This is not always consistent with empirical studies of market data. Later models tend to consider a more general correlation structure, i.e. $\mathrm{d}W_t\mathrm{d}Z_t = \rho\mathrm{d}t$, $\rho \in [-1, 1]$. A relatively recent paper by Maghsoodi [April 2007] studied the Hull and White [1987] stochastic volatility model with general correlation structure, and investigated the exact solution of the model.

Since the Hull-White model only employs the Log-Normal SDE for the underlying, the smile shapes that it produces are rather restricted. The fact that a

general form of local volatility function $\varphi(\cdot)$ leads to a variety of skew/smile shapes encourages us to keep this feature even in stochastic volatility models. A famous example is the SABR model proposed in Hagan et al. [2002]. This is a combination of the CEV model equipped with the Hull and White [1987] model's volatility process. Since its publication, the SABR model has now become one of the standard stochastic volatility models in many different markets. We draw our attention to the SABR model in Chapter 5.

Other SDE formulations for the volatility process have also been considered apart from the Log-Normal type in the Hull and White [1987] model. For example, Scott [1987], and Heston [1993] incorporate mean reversion to the volatility process. See also Fouque et al. [2000a] and Fouque et al. [2000b] for more developments on the mean reverting stochastic volatility models. In Chapter 6, we address to some extent why mean reversion should be considered and study a specific example of the SABR model with mean reverting volatility.

4.4 Desirable properties of a "good" model in practice

We end this chapter by sharing a few thoughts on modelling requirements and discuss the aims that we hope to achieve in the coming chapters.

Up to now, we have reviewed some existing attempts to model the volatility smile. The key questions that we have in mind in the next chapters involve the followings.

- *Which properties are useful for a model in a chosen market?*

  This is not an easy question as modelling requirements vary from market to market and it depends on what we use the model for. The points we want to stress here is that there is no universal model for all applications. Instead of proposing a complicated model that has all useful features, we prefer to work with a simple model that is "nice" enough for our given objectives. This is why in later chapters, we consider the SABR model which is widely used in practice and a modification introducing mean reverting volatility.

- *What can we do given a certain model formulation? Can we compute prices in a timely manner or does the model's flexibility always come with a cost?*

  Having chosen a model that is theoretically suitable for the market and our applications, the next important issue is the implementation. In fact, this problem could be a lot more challenging than the first and very often in practice the rejection of a model comes from its intractability. If there is no available efficient numerical scheme, one will have to tweak the model somehow for practical use in the trading environment. This action does not always leave

the original model to retain its desirable features. In Chapters 5 and 6, we focus on the efficient numerical implementation of the SABR and SABR with mean reversion models. While we want to use the SDE formulations that have nice and suitable properties, we always bear in mind that they must also be implemented in an efficient manner.

- *Given that different stochastic volatility models can achieve a given objective, will they continue to give similar results if we put them in action for other purposes?*

  This is an extension of the first question and we address it in both Chapters 6 and 7. The purpose is to test further the appropriateness of our chosen models. Suppose that there are two available stochastic volatility models that we believe can do the same job (pricing European options and calibrating to market smiles) equally well with comparable computational speed, we are interested in whether they will produce similar results if they are put in a different battle.

  An interesting angle to explore is the hedging problem that is similar to what we have covered in detail in Part I (but with deterministic volatility setting). We want to know how to hedge European options within stochastic volatility models and whether the models yield desirable hedging properties (Chapter 7). Another application that we have in mind is the pricing of other derivatives that are exposed to other information in the market, e.g. future yield curve, forward volatility (Chapter 6). Will different models give different prices?

Chapter 5

On the approximation of the SABR model: a probabilistic approach

In this chapter, we derive a probabilistic approximation for three different versions of the SABR model: Normal, Log-Normal and a displaced diffusion version for the general constant elasticity of variance case. Specifically, we focus on capturing the terminal distribution of the underlying process (conditional on the terminal volatility) to arrive at the implied volatilities of the corresponding European options for all strikes and maturities. Our resulting method allows us to work with a variety of parameters which cover long dated options and highly stress market condition. This is a different feature from other current approaches which rely on the assumption of very small total volatility and usually fail for longer than 10 years maturity or large volatility of volatility.

5.1 Introduction

As noted in Chapter 4, in financial markets we usually observe that implied volatility as a function of strike displays skews (negative slope) or smile shapes. The existence of smiles/skews suggests that the Log-Normal assumption of the underlying process (Black and Scholes [1973]) should be relaxed to develop a more general class of models. In the literature, we have the class of one factor models such as the local volatility models which assume the dependence of volatility on both time and underlying or the more ambitious two factor stochastic volatility models assigning a separate stochastic component to the volatility. We recall from Chapter 4 that although any given market smile and skew can be fitted quite well with the local volatility models, Hagan et al. [2002] pointed out their poor dynamics that predict

wrong movements of the smiles as the underlying moves. This fact implies that even simple derivatives can only be hedged properly with the stochastic volatility models.

We will study one of the most frequently used stochastic volatility models in practice: the SABR model that was originally proposed in Hagan et al. [2002]. It is widely used to model the forward price of the stock or the forward LIBOR/swap rates in the fixed income market. The model is essentially a stochastic volatility extension of the constant elasticity of variance (CEV) model (studied in Schroder [1989] and Cox [1996]) with a lognormal specification of the volatility process. In Hagan et al. [2002], the authors use singular perturbation techniques to obtain explicit, closed-form algebraic formulae for the implied volatility enabling very efficient implementation of the model on a daily basis. The quality of this so-called SABR formula is quite satisfactory given short maturity and strikes not so far from the current underlying. It becomes much poorer for pricing the long dated options or strikes on the wing. In addition, the formula itself has an internal flaw, i.e. implied volatilities for long maturity computed by this formula usually imply negative density of the underlying at very low strike.

A number of other approaches have been developed in the current literature to improve the approximation of the SABR model. Two common techniques are singular perturbation (e.g. Hagan et al. [2002], Hagan et al. [2005] and Wu [2010]) and heat kernel expansion (e.g. Henry-Labordere [2005] and Paulot [2009]). Our method, which is based on a probabilistic framework, focuses on the marginal distribution of the underlying at maturity to arrive at the required implied volatilities. Once we fit an appropriate approximation to the underlying's marginal distribution, implied volatilities can be immediately recovered by inverting the option prices and we do not have the problem regarding negative density as in Hagan et al. [2002].

While the idea is conceptually clear, developing an effective framework for it is not straightforward. One reason is from the solution of the SDE for the underlying process. For the Normal and Log-Normal versions of the SABR model where we are able to write the explicit solutions in distribution for the SDE, the correlation parameter causes the presence of both terminal volatility and realized variance leading to a challenging high dimensional problem. Some authors, hence, assume zero correlation to remove this difficulty and then only the realized variance needs to be considered. This assumption, however, gives rise to a much more restricted SABR model. We keep the general correlation structure but build up our approximation by conditioning the underlying's distribution on the terminal volatility and approximating this distribution. The resulting approximate conditional distribution (with correct mean and variance) has to be theoretically appealing (close to the true distribution) but simple enough to allow for computational efficiency. We propose the Normal and Normal Inverse Gaussian distributions for such purposes.

Another challenge for our approach is the CEV structure of the SABR model which admits no explicit solution. In order to find a way around this, we study the simpler displaced diffusion (DD) model where our previously mentioned method can be applied. The DD models (first studied in Rubinstein [1983]) are the simplest way of incorporating skews even without stochastic volatility in finance literature. Despite the difference between the two models' dynamics, Marris [1999] noted that for a certain model parametrization the option prices and implied volatilities produced by the deterministic CEV and DD models are almost identical across a wide range of strikes and maturities. The comparison is studied further in Svoboda-Greenwood [2009]. See also Rebonato [2002] for a discussion on the CEV and DD models for the interest rate area. Other authors, thereby, adopt the more tractable DD structure with the intuition based on the CEV in the stochastic volatility setting without having investigated the connection between them, e.g. Joshi and Rebonato [2003], Piterbarg [2005] and Larsson [2010]. In this chapter, we attempt to fill in this gap in literature at least numerically with the aim of transferring the intuition from the CEV to DD version of the SABR model for which one can derive an approximation with much less effort.

The chapter is organized as follows. Section 5.2 compares the SABR model and its displaced diffusion version with the mapping connecting them numerically. We develop our approximation in Section 5.3 where we quote the appropriate formulae and match the parameters for implementation. In Section 5.4, we numerically investigate the quality of our approximation in conjunction with other approximations and Monte Carlo simulations. Section 5.5 concludes the chapter.

5.2 SABR model

Under the SABR model, the dynamics of the underlying asset is given by:

$$\begin{aligned} \mathrm{d}F_t &= \sigma_t F_t^{\beta} \mathrm{d}W_t \qquad \beta \in [0,1], \\ \mathrm{d}\sigma_t &= \nu \sigma_t \mathrm{d}Z_t \qquad \nu > 0, \end{aligned} \tag{5.1}$$

where W_t and Z_t are correlated Brownian motions such that $\mathrm{d}W_t \mathrm{d}Z_t = \rho \mathrm{d}t$ for all $t \leq T$ with $\rho \in [-1, 1]$. The model assumes that the underlying process is already a (local) martingale[1] under some equivalent martingale measure.

Each parameter in the SABR model has a specific role in determining the shapes of the skews and smiles. Hagan et al. [2002] was the first to point out these roles through their SABR formula which will be introduced in Section 5.3.2.2. The parameter β has a primary effect on the skew, i.e. reducing β from 1 to 0 gives rise

[1] When $\beta = 1$ and $\rho > 0$, the underlying process is not a martingale.

to more negative (downward) slope of the implied volatility curves. Furthermore, Hagan et al. [2002] also mentioned that β determines the "backbone" which is the curve that the at the money (ATM) volatility traces as F_0 varies. Often, one extracts β from historical data and fixes it upfront for certain markets. It is also noted in Hagan et al. [2002] that market smiles can be fit equally well with any specific value of β. For our later model analysis, we will separate the SABR model into three sub-models:

1. $\beta = 0$: this model is referred to as the Normal SABR model.
2. $\beta = 1$: this model is referred to as the Log-Normal SABR model.
3. $\beta \in (0, 1)$: this model is referred to as the CEV-SABR model.

The ρ-parameter in the SABR model has a similar impact on the skew, i.e. more negative ρ enables a more downward sloping curve. Therefore, ρ is often chosen to match the skew. It also features in general market practice that the implied volatility curves exhibit different levels of curvature. Large curvature usually occurs for short dated options while the smiles tend to flatten out as maturity increases. For that reason, ν known as Volvol (volatility of volatility) is always considered alongside with the market given parameter T (maturity). Finally, the initial volatility σ_0 has a unique role of matching up the ATM implied volatility which corresponds to the most liquid option in any market.

5.2.1 A displaced diffusion version of the SABR model

The non-stochastic CEV model is known to enable a very flexible modelling of volatility skew. Despite this advantage, the CEV structure lacks closed-form solution and numerically it is not very straightforward to implement. The same difficulties also apply to the CEV-SABR model. For our method, we use a much simpler alternative model with the same capability as the CEV-SABR model. In practice, the DD model has been posited for such purpose since it is equally capable of capturing the skews. A further advantage which makes practitioners prefer this model is the fact that the DD structure is very similar to the Log-Normal structure which admits an explicit form for the terminal distribution of the underlying and can be easily handled. Therefore, we study the DD version of the SABR model (DD-SABR) which is specified by the following SDEs

$$\begin{aligned} \mathrm{d}F_t &= \hat{\sigma}_t(F_t + \theta)\mathrm{d}W_t, \\ \mathrm{d}\hat{\sigma}_t &= \nu\hat{\sigma}_t\mathrm{d}Z_t, \\ \mathrm{d}W_t\mathrm{d}Z_t &= \rho\mathrm{d}t. \end{aligned} \tag{5.2}$$

The CEV-SABR and DD-SABR models become comparable via the following mapping

$$\begin{aligned}\hat{\sigma}_t &= \sigma_t \beta F_0^{\beta-1}, \\ \theta &= F_0 \frac{1-\beta}{\beta}.\end{aligned}$$

It is well known that in the deterministic volatility case ($\nu = 0$), the forward dynamics in (5.1) and (5.2) with the above mapping are very similar and implied volatilities produced by the two models are almost identical across a wide range of strikes. This mapping was first discussed in Marris [1999] and studied further in Svoboda-Greenwood [2009]. It was then widely adopted by other authors and practitioners even in the stochastic volatility setting without having been investigated. Note that the mapping is perfect when $\beta = 1$ for which the DD-SABR model collapses to the Log-Normal SABR model. For the rest of the chapter, the DD-SABR model is always equipped with the mapping to match the intended CEV-SABR model.

Having chosen to work with the DD-SABR model, we want to stress the importance of the CEV-SABR one and compare the two models numerically for completeness. We will split our comparison into two parts. The first part is about the mapping quality when strikes are near the money while the second one focuses on the wing behaviour. The reason is that the SABR model best represents the market given strikes not too far from the current underlying level. When pricing long-dated options, both models tend to break down in the wings since each of them has its own shortcomings. In our comparison, we only look at the smiles exhibited by the two models when their ATM volatilities are matched as this is the comparison that matters in practice.

5.2.1.1 Near the money

We have systematically investigated the mapping under different regimes and scenarios when strikes are not far from at the money. In the results presented here, the parameters are taken to be consistent with our later numerical study and representative enough so that similar results are expected to hold for all cases.

Figures 5.1, 5.2 and 5.3 illustrate the effects of both ν and T on the mapping. For up to medium long maturity (15 years) and low Volvol ν (0.3), the mapping is quite accurate with errors recorded to be very small across all strikes. The maximum error is about 60 basis points (bp) at the lowest strike. When Volvol is higher, the DD-SABR model displays more curvature on the smiles but the differences still remain acceptably small (maximum 100 bp). We then take the maturity to be very long (20 and 30 years) with low Volvol ν as usually expected in practice (figure

5.4). The resulting plots show that the mapping starts breaking down as the shapes of two implied volatility curves are not entirely in line with each other. The DD-SABR model produces progressively steeper skews while the CEV-SABR's curves tend to kick up at the right wing, i.e. leading to positive errors for low strikes and negative errors for large strikes. This effect becomes much more significant when we deal with strikes that are far from at the money. The rare cases of high Volvol and long maturity are not presented here but one observes that similar effects hold throughout.

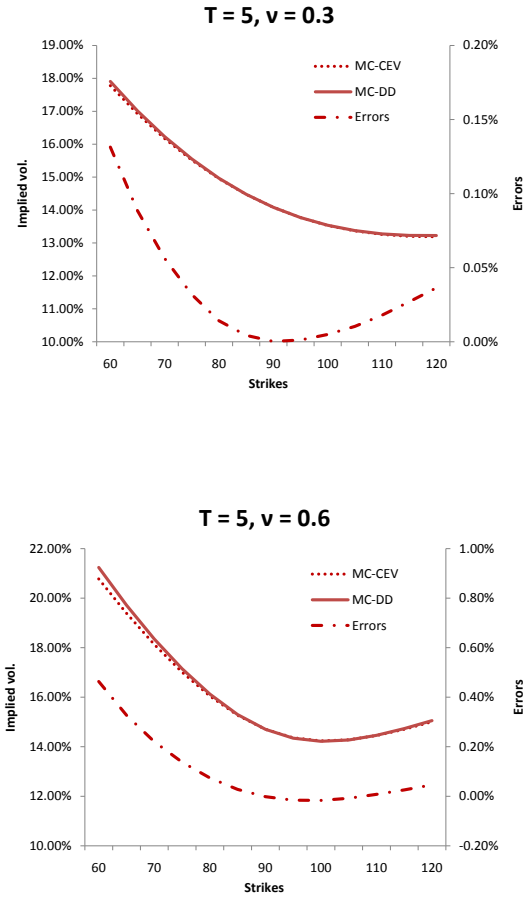


Figure 5.1: Effects of maturity T and Volvol ν on the mapping when the ATM are matched. Parameters: $\beta = 0.5, \rho = -0.2, \sigma_0 = 130\%, F_0 = 90$. MC-CEV: CEV-SABR MC solution, MC-DD: DD-SABR MC solution, Errors: MC-DD minus MC-CEV.

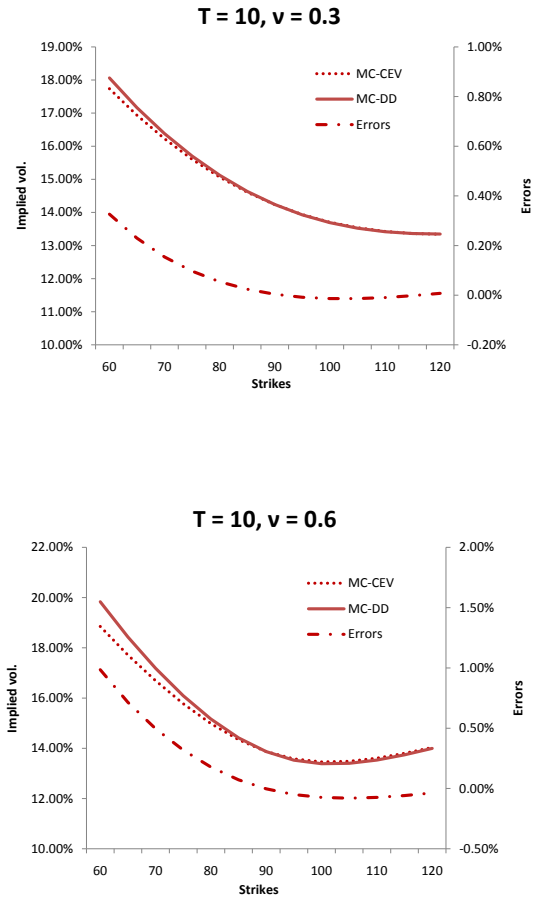


Figure 5.2: Effects of maturity T and Volvol ν on the mapping when the ATM are matched. Parameters: $\beta = 0.5, \rho = -0.2, \sigma_0 = 130\%, F_0 = 90$. MC-CEV: CEV-SABR MC solution, MC-DD: DD-SABR MC solution, Errors: MC-DD minus MC-CEV.

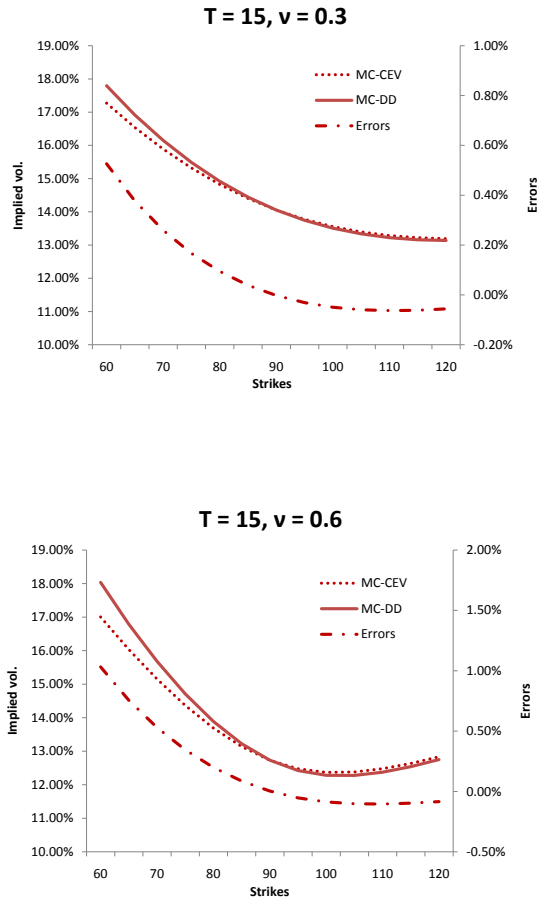


Figure 5.3: Effects of maturity T and Volvol ν on the mapping when the ATM are matched. Parameters: $\beta = 0.5, \rho = -0.2, \sigma_0 = 130\%, F_0 = 90$. MC-CEV: CEV-SABR MC solution, MC-DD: DD-SABR MC solution, Errors: MC-DD minus MC-CEV.

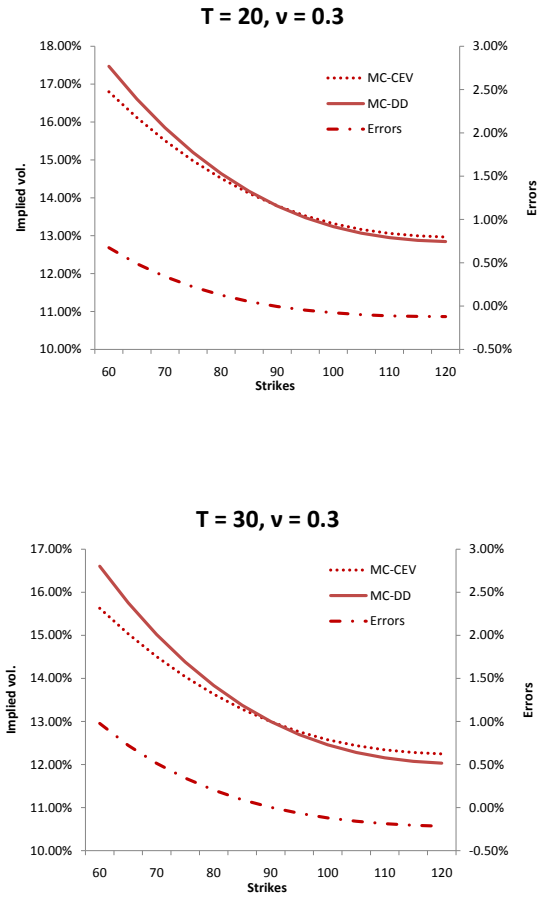


Figure 5.4: Effect of very long maturity T on the mapping when the ATM are matched. Parameters: $\beta = 0.5, \rho = -0.2, \sigma_0 = 130\%, F_0 = 90$.

We mentioned that both ρ and β affect the skew. In figure 5.5, it is seen that the correlation parameter ρ does not really affect the mapping and the error curves look almost identical. On the other hand, β as illustrated in figure 5.6 has a stronger influence and the mapping tends to be less accurate for smaller β. This makes sense since perfect mapping is obtained as β approaches one. For low value of β, the displaced diffusion coefficient θ is large enabling more probability mass to be assigned to negative values of F_T while the absorbing barrier of the CEV structure also plays a more significant role.

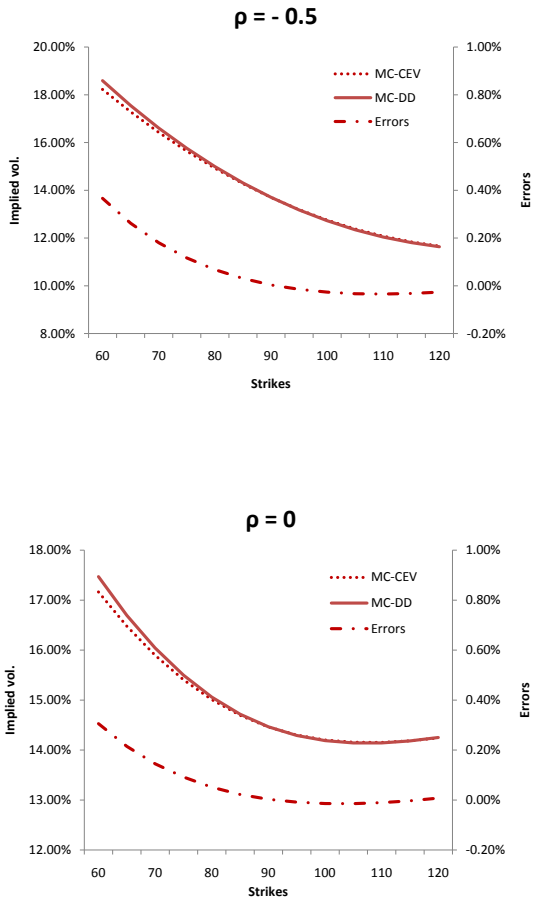


Figure 5.5: Effect of ρ on the mapping when the ATM are matched. Parameters: $\beta = 0.5, T = 10, \nu = 0.3, \sigma_0 = 130\%, F_0 = 90$.

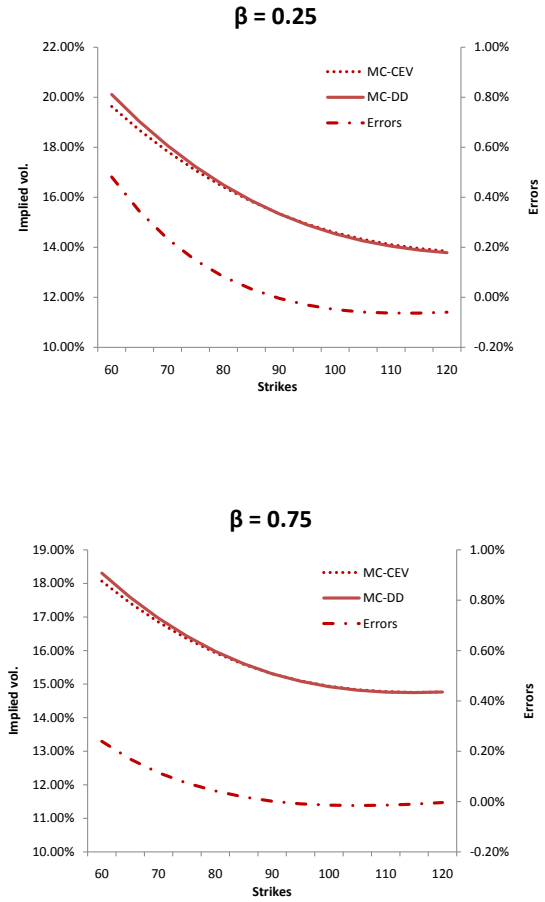


Figure 5.6: Effect of β on the mapping when the ATM are matched. Parameters: $\rho = -0.2, T = 10, \nu = 0.3, F_0 = 90$, σ_0 is chosen for each case so that the ATM are comparable.

5.2.1.2 Implied volatilities in the wings

We have further investigated the behaviour in the wings. In the results presented here, the parameters are taken to represent typical market swaption smiles of different maturities (figures 5.7 and 5.8). We chose to work with swaption data as strikes being far from at the money is observed more often in the interest rate market. While high strikes are not really a problem, the gap between two models gets bigger as the strike gets lower. When the strike is sufficiently low (ATM - 200 bp), the error can approach 3 to 4% which is quite significant in practice. For increasing maturity (20 and 30 years), the mapping completely breaks down for "ATM - 200 bp" strike even with very low ν. This fact was addressed in Svoboda-Greenwood [2009] in detail. The author argues that even in the deterministic volatility setting, the mapping may work well given the assumption that forward interest rates are "not too low" and their percentage volatilities are "reasonable". When such assumption fails, a greater portion of the probability density function is likely to fall

in the negative rates region for the DD process while a large part of the distribution is absorbed at zero for the CEV process over intermediate maturities. These effects become more pronounced for longer maturity. We report these results for the data used in figures 5.7 and 5.8 in table 5.1. For the 30 year maturity case, it is seen that around a quarter of the mass is given to the absorbing barrier and the negative rates region. Therefore, the mapping can no longer be justified. We want to stress that this is not really a problem as both models are not good enough in practice here.

In the next section, we derive an approximation for the models (excluding CEV-SABR). Note that with these models practitioners are only interested in around the ATM region. From that perspective, the approximation is valid for all different asset classes including interest rates.

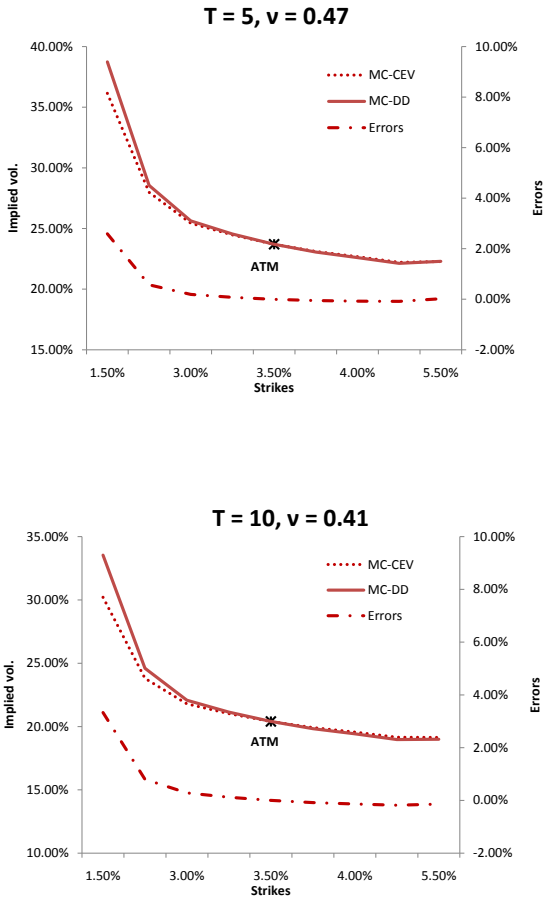


Figure 5.7: Implied volatilities under different models. Parameters: $\beta = 0.5, \rho = -0.2, \sigma_0 = 4.30\%, 3.80\%$ as maturity increases, respectively.

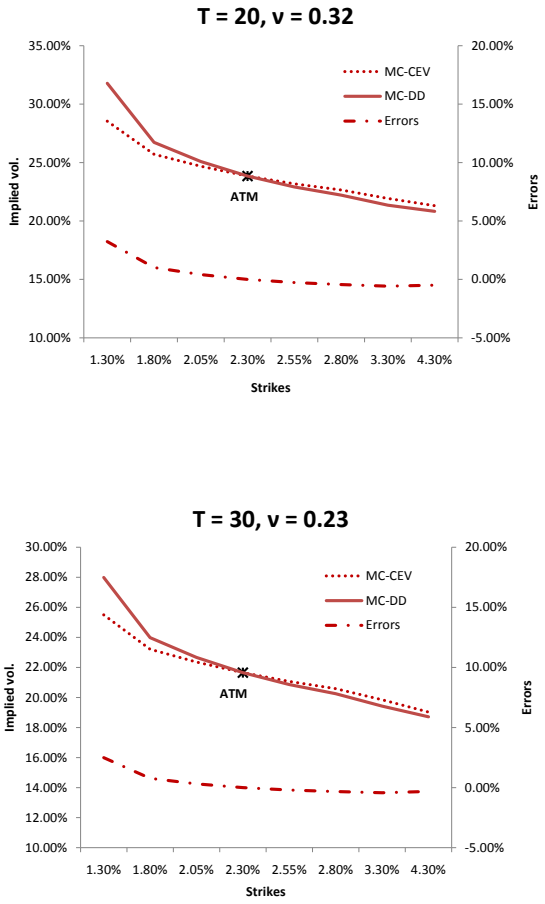


Figure 5.8: Implied volatilities under different models. Parameters: $\beta = 0.5, \rho = -0.2, \sigma_0 = 4.10\%, 3.70\%$ as maturity increases, respectively.

T	5 Y	10 Y	20 Y	30 Y
CEV-SABR	5.18 %	9.77%	25.85%	30.15%
DD-SABR	4.03 %	7.67%	19.47%	22.60%

Table 5.1: Probability mass assigned to the absorbing barrier (CEV-SABR) and the negative rates region (DD-SABR) for the four cases considered in figures 5.7 and 5.8 (computed by direct Monte Carlo simulation).

5.3 A probabilistic approximation

5.3.1 Approximating the terminal distribution

By the fundamental pricing formula and tower property, today's numeraire-rebased price of a vanilla call option struck at some strike K is given by

$$\begin{aligned}\mathbb{C}_0(K, F_0) &= \mathbb{E}\left[(F_T - K)^+\right] \\ &= \mathbb{E}\left[\mathbb{E}\{(F_T - K)^+ | \sigma_T\}\right]. \end{aligned} \tag{5.3}$$

Assuming that we have in mind some distribution for $F_T|\sigma_T$: the conditional distribution of F_T given σ_T (see Sections 5.3.2 and 5.3.3), then the conditional expectation above can be evaluated as a double integral. Recall that σ_T has a known Log-Normal distribution as the SDE it solves has an explicit solution. To keep the notation simple and transparent, we introduce the process s that represents the level of assets and function $g(.)$ to transform it back to the underlying process F, i.e. $F_t = g(s_t)$. As the first stepping stone, we will write down the exact solutions in distribution to the SDE for our reference models (see Appendix 5.A for details).

- **Normal SABR**:

$$\begin{aligned} s_T &\triangleq F_0 + \frac{\rho}{\nu}(\sigma_T - \sigma_0) + \sqrt{1-\rho^2} V_T^{\frac{1}{2}} G, \\ g(s) &= s. \end{aligned} \tag{5.4}$$

- **Log-Normal SABR**:

$$\begin{aligned} s_T &\triangleq \ln F_0 + \frac{\rho}{\nu}(\sigma_T - \sigma_0) - \frac{1}{2} V_T + \sqrt{1-\rho^2} V_T^{\frac{1}{2}} G, \\ g(s) &= e^s. \end{aligned} \tag{5.5}$$

- **DD-SABR**:

$$\begin{aligned} s_T &\triangleq \ln(F_0 + \theta) + \beta F_0^{\beta-1} \frac{\rho}{\nu}(\sigma_T - \sigma_0) - \frac{1}{2}\beta^2 F_0^{2\beta-2} V_T + \beta F_0^{\beta-1}\sqrt{1-\rho^2} V_T^{\frac{1}{2}} G, \\ \theta &= F_0 \frac{1-\beta}{\beta}, \\ g(s) &= e^s - \theta. \end{aligned} \tag{5.6}$$

Here $V_T := \int_0^T \sigma_t^2 \mathrm{d}t$ is the realized variance and G is a standard Normal random variable independent of σ_T and V_T. We aim to approximate the conditional distribution of $s_T|\sigma_T$ by replacing it with some suitable random variable with the same conditional mean and variance. In each case, the realized variance V_T plays a central role in our calculation and analysis so we will treat its moments separately in the

following proposition.

Proposition 1 *Assume that the dynamics of the volatility is governed by a Log-Normal process with Volvol $\nu > 0$, i.e. $\mathrm{d}\sigma_t = \nu\sigma_t \mathrm{d}Z_t$ where Z is a Brownian motion. The first two* ***conditional moments of the realized variance*** *V_T have the following analytical expressions:*

$$\mathbb{E}(V_T|\sigma_T) = \frac{\sigma_0^2\sqrt{T}}{2\nu}\frac{\left[\Phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}}+\nu\sqrt{T}\right)-\Phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}}-\nu\sqrt{T}\right)\right]}{\phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}}+\nu\sqrt{T}\right)}, \quad (5.7)$$

$$\begin{aligned}\mathbb{E}(V_T^2|\sigma_T) = & -\frac{\sigma_0^4\sqrt{T}}{4\nu^3}\left(1+e^{2\ln(\sigma_T/\sigma_0)}\right)\frac{\left[\Phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}}+\nu\sqrt{T}\right)-\Phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}}-\nu\sqrt{T}\right)\right]}{\phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}}+\nu\sqrt{T}\right)} \\ & +\frac{\sigma_0^4\sqrt{T}}{4\nu^3}\frac{\left[\Phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}}+2\nu\sqrt{T}\right)-\Phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}}-2\nu\sqrt{T}\right)\right]}{\phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}}+2\nu\sqrt{T}\right)}, \quad (5.8)\end{aligned}$$

where $\phi(.)$ and $\Phi(.)$ are the Normal density and cumulative distribution functions respectively.

Proof: Appendix 5.B.

5.3.2 Normal approximation

We first consider the Normal distribution for the approximation of $s_T|\sigma_T$ as it appears to be very tractable and efficient to use in practice. Another motivation for choosing the Normal distribution comes from an earlier numerical investigation in Mitra [2010]. In this work, the conditional distribution of $s_T|\sigma_T$ was seen to be quite close to Normal through examination of the Q-Q plots (see Section 5.3.3 for further discussion). In order to implement this approximation, we first need to calculate the exact conditional mean and variance of s_T

$$\begin{aligned}\mu(\sigma_T) &= \mathbb{E}(s_T|\sigma_T), \\ \eta^2(\sigma_T) &= \mathrm{Var}(s_T|\sigma_T),\end{aligned}$$

and then replace the conditional distribution of $s_T|\sigma_T$ by a Normal random variable with mean $\mu(\sigma_T)$ and variance $\eta^2(\sigma_T)$. One will then be able to calculate the call option prices by (5.3) and obtain the implied volatilities. The analytical formulae for $\mu(\sigma_T)$ and $\eta^2(\sigma_T)$ are quoted in the following proposition.

Proposition 2 *: The conditional mean and variance of s_T for the reference models are given by the following closed-form expressions:*

- ***Normal SABR**:*

$$\begin{aligned}\mu(\sigma_T) &= F_0 + \frac{\rho}{\nu}(\sigma_T - \sigma_0),\\ \eta^2(\sigma_T) &= (1-\rho^2)\frac{\sigma_0^2\sqrt{T}}{2\nu}\frac{\left[\Phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}} + \nu\sqrt{T}\right) - \Phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}} - \nu\sqrt{T}\right)\right]}{\phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}} + \nu\sqrt{T}\right)}.\end{aligned} \tag{5.9}$$

- ***Log-Normal SABR** and **DD-SABR**:*

$$\begin{aligned}&\mu(\hat{\sigma}_T)\\ =\ & \ln(F_0+\theta) + \frac{\rho}{\nu}(\hat{\sigma}_T - \hat{\sigma}_0) - \frac{\hat{\sigma}_0^2\sqrt{T}}{4\nu}\frac{\left[\Phi\left(\frac{\ln(\hat{\sigma}_T/\hat{\sigma}_0)}{\nu\sqrt{T}} + \nu\sqrt{T}\right) - \Phi\left(\frac{\ln(\hat{\sigma}_T/\hat{\sigma}_0)}{\nu\sqrt{T}} - \nu\sqrt{T}\right)\right]}{\phi\left(\frac{\ln(\hat{\sigma}_T/\hat{\sigma}_0)}{\nu\sqrt{T}} + \nu\sqrt{T}\right)},\\ &\eta^2(\hat{\sigma}_T)\\ =\ & \frac{\hat{\sigma}_0^2\sqrt{T}}{2\nu}\left((1-\rho^2) - \frac{(\hat{\sigma}_T^2+\hat{\sigma}_0^2)}{8\nu^2}\right)\frac{\left[\Phi\left(\frac{\ln(\hat{\sigma}_T/\hat{\sigma}_0)}{\nu\sqrt{T}} + \nu\sqrt{T}\right) - \Phi\left(\frac{\ln(\hat{\sigma}_T/\hat{\sigma}_0)}{\nu\sqrt{T}} - \nu\sqrt{T}\right)\right]}{\phi\left(\frac{\ln(\hat{\sigma}_T/\hat{\sigma}_0)}{\nu\sqrt{T}} + \nu\sqrt{T}\right)}\\ &+\frac{\hat{\sigma}_0^4\sqrt{T}}{16\nu^3}\frac{\left[\Phi\left(\frac{\ln(\hat{\sigma}_T/\hat{\sigma}_0)}{\nu\sqrt{T}} + 2\nu\sqrt{T}\right) - \Phi\left(\frac{\ln(\hat{\sigma}_T/\hat{\sigma}_0)}{\nu\sqrt{T}} - 2\nu\sqrt{T}\right)\right]}{\phi\left(\frac{\ln(\hat{\sigma}_T/\hat{\sigma}_0)}{\nu\sqrt{T}} + 2\nu\sqrt{T}\right)}\\ &-\frac{\hat{\sigma}_0^4 T}{16\nu^2}\left(\frac{\left[\Phi\left(\frac{\ln(\hat{\sigma}_T/\hat{\sigma}_0)}{\nu\sqrt{T}} + \nu\sqrt{T}\right) - \Phi\left(\frac{\ln(\hat{\sigma}_T/\hat{\sigma}_0)}{\nu\sqrt{T}} - \nu\sqrt{T}\right)\right]}{\phi\left(\frac{\ln(\hat{\sigma}_T/\hat{\sigma}_0)}{\nu\sqrt{T}} + \nu\sqrt{T}\right)}\right)^2,\end{aligned} \tag{5.10}$$

where

$$\begin{aligned}\theta &= F_0\frac{1-\beta}{\beta},\\ \hat{\sigma}_t &= \sigma_t\beta F_0^{\beta-1}.\end{aligned}$$

Proof: By Proposition 1 and direct calculations (see Appendix 5.C for more details). Clearly, the formulae in (5.10) for the Log-Normal SABR model are obtained when $\beta = 1$.

5.3.2.1 Implementation: advantages and disadvantages

We apply the formulae derived in the last section to the direct calculations of vanilla call option prices for all strikes. Since there is a one to one correspondence between

the volatility process σ and its driving Brownian motion Z (through the SDE of σ)

$$\begin{aligned} Z_T &= \frac{\ln(\sigma_T/\sigma_0) + \frac{1}{2}\nu^2 T}{\nu}, \\ Z_T &\sim \mathcal{N}(0,T), \end{aligned}$$

we can also express the conditional mean and variance in terms of Z_T. Consequently, the inner conditional expectation in (5.3) has the equivalent expression $\mathbb{E}\left[(F_T-K)^+|Z_T\right]$ and (5.3) now reads

$$\mathbb{C}_0(K,F_0) = \int_{-\infty}^{\infty} \mathbb{E}[(g(s_T)-K)^+|Z_T=x] f_{Z_T}(x)\mathrm{d}x,$$

where $g(.)$ is the appropriate transformation for the chosen β and $f_{Z_T}(x) = e^{-\frac{x^2}{2T}}/\sqrt{2\pi T}$ is the probability density function of Z_T. After some direct calculations we obtain:

1. **Normal SABR**:

$$\begin{aligned} &\mathbb{C}_0(K,F_0) \\ =\ & \int_{-\infty}^{\infty} \left[\sqrt{\eta^2(x)}\phi\left(\frac{K-\mu(x)}{\sqrt{\eta^2(x)}}\right) + (\mu(x)-K)\left(1-\Phi\left(\frac{K-\mu(x)}{\sqrt{\eta^2(x)}}\right)\right)\right] \frac{e^{\frac{-x^2}{2T}}}{\sqrt{2\pi T}}\mathrm{d}x. \end{aligned} \tag{5.11}$$

2. **Log-Normal SABR**:

$$\begin{aligned} &\mathbb{C}_0(K,F_0) \\ =\ & \int_{-\infty}^{\infty} \left[e^{\mu(x)+\frac{\eta^2(x)}{2}}\Phi\left(\frac{\mu(x)+\eta^2(x)-\ln K}{\sqrt{\eta^2(x)}}\right) - K\Phi\left(\frac{\mu(x)-\ln K}{\sqrt{\eta^2(x)}}\right)\right] \frac{e^{\frac{-x^2}{2T}}}{\sqrt{2\pi T}}\mathrm{d}x. \end{aligned} \tag{5.12}$$

Remark 8 *: For the DD-SABR model, we have exactly the same formula as* (5.12) *with K replaced by $K+\theta$.*

Both (5.11) and (5.12) are simple one-dimensional integrals and can therefore be evaluated easily by some efficient numerical routine. We want to emphasize this point because we think it is crucial. Although the Normal approximation, as we shall see later, does not appear to be the best choice theoretically, it is the only one that could compete with other asymptotic approximations in terms of computational time and this is an important consideration for any practical model. Consequently, one should always look at the regimes when it works well and not so well. Despite its convenience and simple form, the Normal approximation admits a potential numerical problem as described in the following remark.

Remark 9 *: For both the Log-Normal SABR and DD-SABR models,* $\mathbb{E}(V_T^2|\sigma_T)$ *and hence* $\eta^2(\sigma_T)$ *become very large when* $\nu^2 T$ *is large can be observed from equation* (5.8)*. For certain parameter choices, the growth rate of* $\mathbb{E}[(g(s_T)-K)^+|Z_T=x]$ *in equation* (5.12) *is not balanced by the rate of decay of* $f_{Z_T}(x)$ *and hence, leads to the numerical divergence of the integral. This problem can be illustrated by the following figure*

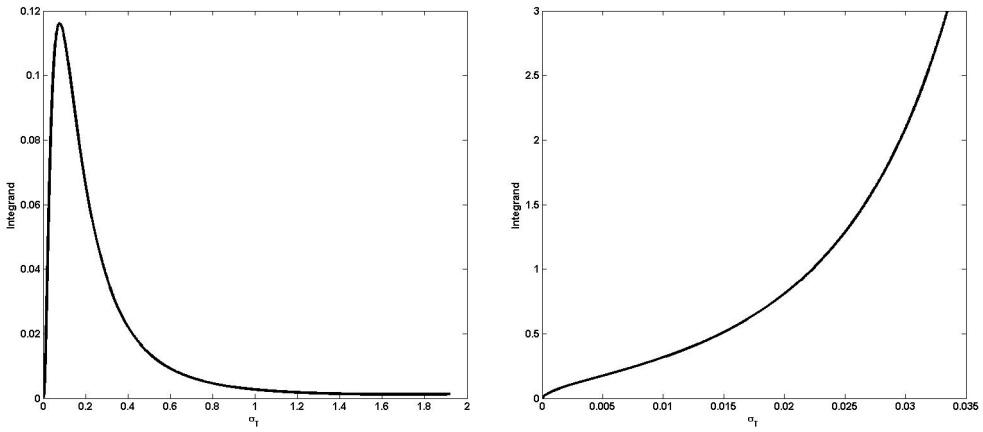


Figure 5.9: The integrand of (5.12) as a function of σ_T. Left plot: $\beta=1, \rho=-0.5, F_0=90, K=90, T=10, \nu=0.3, \sigma_0=15\%$, right plot: $\beta=1, \rho=-0.5, F_0=90, K=90, T=15, \nu=0.6, \sigma_0=15\%$.

As a result, prices can not be calculated correctly when the numerical convergence fails. In principle, one can do the following

$$\begin{aligned}\mathbb{C}_0(K,F_0) &= \int_{-\infty}^{\infty}\mathbb{E}[(g(s_T)-K)^+|Z_T=x]\frac{e^{\frac{-x^2}{2T}}}{\sqrt{2\pi T}}\mathrm{d}x \\ &\approx \int_{\underline{z}}^{\overline{z}}\mathbb{E}[(g(s_T)-K)^+|Z_T=x]\frac{e^{\frac{-x^2}{2T}}}{\sqrt{2\pi T}}\mathrm{d}x,\end{aligned}$$

where $\underline{z}$ *and* $\overline{z}$ *are the appropriate lower and upper limits for the numerical integration. For some regimes of large* $\nu^2 T$*,* $\overline{z}$ *cannot be chosen to give the numerical convergence. In practice, one can truncate the integral at a much lower* $\overline{z}$ *to avoid this issue as the volatility process is unlikely to hit a very high level at maturity. If the truncated value is too low, the density function will have to be re-normalized,*

that is

$$\begin{aligned}\mathbb{C}_0(K,F_0) &= \int_{\underline{z}}^{\overline{z}} \mathbb{E}[(g(s_T)-K)^+|Z_T=x]\tilde{f}_{Z_T}(x)\mathrm{d}x,\\ \tilde{f}_{Z_T}(x) &= \frac{e^{\frac{-x^2}{2T}}}{\int_{\underline{z}}^{\overline{z}} e^{\frac{-u^2}{2T}}\mathrm{d}u}.\end{aligned}$$

5.3.2.2 A comparison with other approximations

We briefly review some attempts made by other authors to approximate the SABR model and compare them with the Normal approximation in terms of implied volatility. We want to single out those that have already been tested numerically. Table 5.2 gives a brief overview of various approximations labeled by authors for all sub-models.

Authors	Normal SABR	Log-Normal SABR	CEV-SABR	DD-SABR
Hagan et al. [2002]	tested	tested	tested	not tested
Obloj [2008]	tested	tested	tested	not available
Paulot [2009]	not tested	not tested	tested	not available
Johnson and Nonas [2009]	not available	tested	not tested	not available
Wu [2010]	tested	tested	tested	not available
Larsson [2010]	tested	tested	not available	tested

Table 5.2: Checklist of the most current approximations for the SABR model.

The SABR formula in Hagan et al. [2002] is the original and, perhaps, the most popular amongst the listed works in this table owing to its algebraic closed-form expression. Henceforth, we take the SABR formula as the benchmark approximation for our comparison. In the SABR formula, the Black implied volatility $\sigma_B(K,F_0)$ for a vanilla call (or put) option written on the forward price S struck at some strike K has the following form

$$\begin{aligned}\sigma_B(K,F_0) &= \frac{\sigma_0}{(F_0K)^{(1-\beta)/2}\left\{1+\frac{(1-\beta)^2}{24}\ln^2\frac{F_0}{K}+\frac{(1-\beta)^4}{1920}\ln^4\frac{F_0}{K}+...\right\}}\left(\frac{z}{x(z)}\right)\\ &\quad\left\{1+\left[\frac{(1-\beta)^2}{24}\frac{\sigma_0^2}{(F_0K)^{1-\beta}}+\frac{1}{4}\frac{\rho\beta\sigma_0\nu}{(F_0K)^{\frac{1-\beta}{2}}}+\frac{2-3\rho^2}{24}\nu^2\right]T+...\right\},\end{aligned} \tag{5.13}$$

where

$$\begin{aligned} z &= \frac{\nu}{\sigma_0}(F_0K)^{(1-\beta)/2}\ln\frac{F_0}{K}, \\ x(z) &= \ln\left\{\frac{\sqrt{1-2\rho z+z^2}+z-\rho}{1-\rho}\right\}. \end{aligned} \tag{5.14}$$

The ATM Black implied volatility reduces to

$$\sigma_B(F_0,F_0) = \sigma_0 F_0^{\beta-1}\left\{1+\left[\frac{(1-\beta)^2}{24}\frac{\sigma_0^2}{F_0^{2-2\beta}}+\frac{1}{4}\frac{\rho\beta\sigma_0\nu}{F_0^{1-\beta}}+\frac{2-3\rho^2}{24}\nu^2\right]T+...\right\}. \tag{5.15}$$

We borrow the same technique[2] to derive an equivalent Black implied volatility formula for the DD-SABR model (see Appendix 5.D)

$$\begin{aligned} \sigma_B(K,F_0) &= \hat{\sigma}_0\frac{\sqrt{(F_0+\theta)(K+\theta)}}{\sqrt{F_0K}}\left(\frac{1+\frac{1}{24}\ln^2\frac{F_0+\theta}{K+\theta}+\frac{1}{1920}\ln^4\frac{F_0+\theta}{K+\theta}+\dots}{1+\frac{1}{24}\ln^2\frac{F_0}{K}+\frac{1}{1920}\ln^4\frac{F_0}{K}+\dots}\right)\left(\frac{z}{x(z)}\right) \\ &\quad \left\{1+\left[\frac{2\theta/\sqrt{F_0K}+\theta^2/(F_0K)}{24}\hat{\sigma}_0^2+\frac{1}{4}\rho\nu\hat{\sigma}_0+\frac{2-3\rho^2}{24}\nu^2\right]T+\dots\right\}, \end{aligned} \tag{5.16}$$

where

$$\begin{aligned} z &= \frac{\nu}{\hat{\sigma}_0}\ln\frac{F_0+\theta}{K+\theta}, \\ \theta &= F_0\frac{1-\beta}{\beta}, \\ \hat{\sigma}_0 &= \sigma_0\beta F_0^{\beta-1}, \end{aligned}$$

and $x(z)$ has the same form as (5.14). For the special case of the ATM option, the formula reduces to

$$\begin{aligned} & \sigma_B(F_0,F_0) \\ = \;& \hat{\sigma}_0\frac{F_0+\theta}{F_0}\left\{1+\left[\frac{2\theta/F_0+\theta^2/F_0^2}{24}\hat{\sigma}_0^2+\frac{1}{4}\rho\nu\hat{\sigma}_0+\frac{2-3\rho^2}{24}\nu^2\right]T+\dots\right\} \\ = \;& \sigma_0F_0^{\beta-1}\left\{1+\left[\frac{2\frac{1-\beta}{\beta}+\frac{(1-\beta)^2}{\beta^2}}{24}\sigma_0^2\beta^2F_0^{2\beta-2}+\frac{1}{4}\rho\nu\sigma_0\beta F_0^{\beta-1}+\frac{2-3\rho^2}{24}\nu^2\right]T+\dots\right\} \\ = \;& \sigma_0F_0^{\beta-1}\left\{1+\left[\frac{1-\beta^2}{24}\frac{\sigma_0^2}{F_0^{2-2\beta}}+\frac{1}{4}\frac{\rho\beta\sigma_0\nu}{F_0^{1-\beta}}+\frac{2-3\rho^2}{24}\nu^2\right]T+\dots\right\}. \end{aligned} \tag{5.17}$$

For the rest of the chapter, we will refer to (5.16) and (5.17) as the DD-SABR for-

[2]We take into account the main criticism of the SABR formula pointed out in Obloj [2008] whilst deriving this formula.

mula. One can immediately recognize a lot of similarities between this formula and the SABR formula given strikes near the money and short maturity. A systematic comparison of the Normal approximation with the SABR and DD-SABR formulae will be addressed in Section 5.4. Meanwhile, we summarize the results of other established approximations in conjunction with the SABR formula and emphasize the Normal approximation's superiority.

Most of the approximations listed in table 5.2 fail or lose their precision when $T > 10$ years even with low ν, e.g. both Wu [2010] and Larsson [2010] focus on maturity less than 5 years or Paulot [2009] completely breaks down for $\nu^2 T > 1.6$. The reason is that most of the techniques (singular perturbation or heat kernel expansion) are based on the assumption of small total volatility[3] $\nu^2 T$ to allow for accurate asymptotic expansions up to the second order. As discussed in Section 5.3.2.1, the total volatility $\nu^2 T$ also affects the Normal approximation to some extent. An intuitive reason for this adverse effect is that a larger value of $\nu^2 T$ will push the true conditional distribution of $s_T|\sigma_T$ further away from Normal. However, in the results presented in Section 5.4, the Normal approximation is shown to perform quite well for the Normal SABR model up to 30 years maturity or very large $\nu^2 T \approx 10.8$. For the other sub-models, it works well up to 15 years maturity or $\nu^2 T \approx 1.8$. A further advantage of the Normal approximation over the current approaches is that it always yields a proper density function for the underlying while the other techniques sometimes result in negative density at the low strike region for long maturity, e.g. the SABR formula. This issue is addressed in Obloj [2008] and Johnson and Nonas [2009] but the problem still remains.

5.3.3 Normal Inverse Gaussian approximation

As hinted previously, the true conditional distribution of $s_T|\sigma_T$ can be far from Normal for some parameter sets. We track down this flaw by looking at the Q-Q plots of the standardized conditional sample of $s_T|\sigma_T$ against the standard Normal distribution. In figure 5.10, the results show that even when the Normal approximation works, the true conditional distribution displays much heavier tails than the Normal distribution. We even observe more left skewness as σ_T gets bigger.

[3]Other authors usually use $\epsilon = \nu\sqrt{T}$ as the perturbation parameter. Theoretically, they require this parameter to be much smaller than 1 to give precise results, e.g. Hagan et al. [2002], Hagan et al. [2005] and Wu [2010]. In practice, such requirement can only be satisfied for very short maturity (less than 10 years).

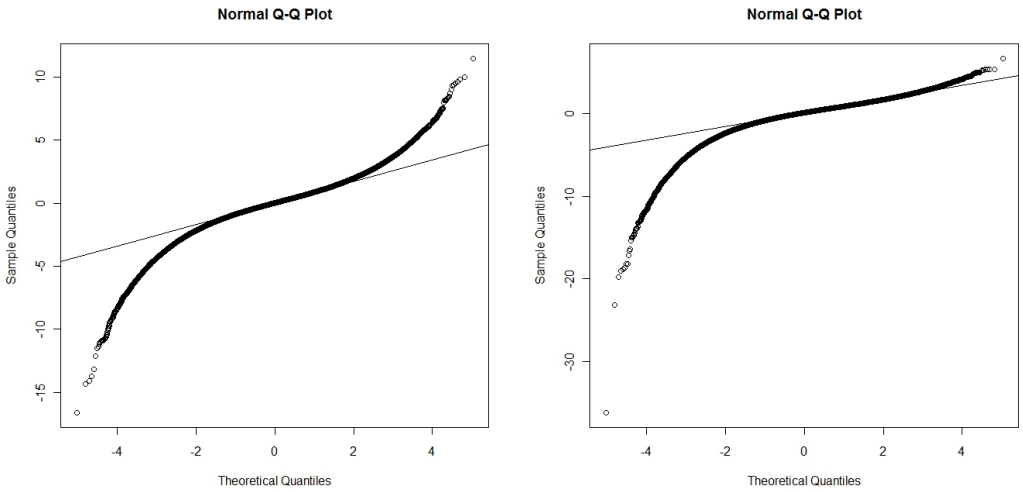


Figure 5.10: Normal Q-Q plots: standardized conditional samples of $s_T|\sigma_T$ against the standard Normal distribution. Common parameters: $\beta = 1, \rho = -0.5, F_0 = 90, \sigma_0 = 5\%$. left plot: $T = 15, \nu = 0.3, \sigma_T = 5\%$, right plot: $T = 15, \nu = 0.3, \sigma_T = 50\%$.

The breakdown of the Normal approximation for certain parameter choices leads us to a further investigation of a more flexible distribution which can capture the skewness and heavy tails. We propose the Normal Inverse Gaussian (NIG) distribution for such purpose. NIG is quite popular in finance, especially in the financial econometrics literature, for instance Barndorff-Nielsen [1997].

Under the NIG approximation, we assume

$$s_T|\sigma_T \sim \mathcal{NIG}(\hat{\alpha}, \hat{\beta}, \hat{\mu}, \hat{\delta}),$$

where the parameters are to be chosen. The NIG density function is defined as follows

$$f_{NIG}(s; \hat{\alpha}, \hat{\beta}, \hat{\mu}, \hat{\delta}) = \frac{\hat{\alpha}}{\hat{\delta}} \exp(\hat{\delta}\sqrt{\hat{\alpha}^2 - \hat{\beta}^2} - \hat{\beta}\hat{\mu}) \frac{K_1\left(\hat{\alpha}\hat{\delta}\sqrt{1 + (\frac{s-\hat{\mu}}{\hat{\delta}})^2}\right)}{\sqrt{1 + (\frac{s-\hat{\mu}}{\hat{\delta}})^2}} \exp(\hat{\beta}s), \quad (5.18)$$

where $s \in \mathbb{R}$ and each parameter has a specific role: $\hat{\alpha} > 0$ determines the tail heaviness of the distribution, $\hat{\delta} > 0$ is the scale parameter, $\hat{\mu} \in \mathbb{R}$ is the location parameter, and $|\hat{\beta}| < \hat{\alpha}$ controls the asymmetry of the distribution. The function $K_1(.)$ is the modified Bessel function of the third kind with index 1. The Gaussian distribution is obtained as $\hat{\alpha} \to \infty$. Despite the involvement of a number of free parameters, the process of matching them to the intended distribution is actually very straightforward as we shall see later.

5.3.3.1 Matching Parameters

We now describe an efficient way to match the NIG parameters. We use the fact that a NIG random variable X can be expressed as the Normal variance-mean mixture form:

$$X = \hat{\mu} + \hat{\beta}Y + \sqrt{Y}G, \tag{5.19}$$

where the mixing random variable Y follows an Inverse Gaussian (IG) distribution (see Barndorff-Nielsen [1997]) and G is a standard Normal random variable that is independent of Y. If

$$\begin{aligned} Y &\sim \mathcal{IG}(\hat{\delta}, \sqrt{\hat{\alpha}^2 - \hat{\beta}^2}), \\ \mathbb{E}(Y) &= \frac{\hat{\delta}}{\sqrt{\hat{\alpha}^2 - \hat{\beta}^2}}, \\ \mathrm{Var}(Y) &= \frac{\hat{\delta}}{\left(\sqrt{\hat{\alpha}^2 - \hat{\beta}^2}\right)^3}, \end{aligned}$$

then

$$X \sim \mathcal{NIG}(\hat{\alpha}, \hat{\beta}, \hat{\mu}, \hat{\delta}).$$

Therefore, matching the mean and variance of the mixing random variable is adequate to capture those of the corresponding NIG random variable. It is clear from (5.4) to (5.6) that conditioned on σ_T, s_T will have a similar form as (5.19). We will now express the NIG parameters in terms of σ_T.

- For $\beta = 0$: the mixing random variable is $(1-\rho^2)V_T|\sigma_T$. We first match the location and asymmetry parameters

$$\begin{aligned} \hat{\mu}(\sigma_T) &= F_0 + \frac{\rho}{\nu}(\sigma_T - \sigma_0), \\ \hat{\beta}(\sigma_T) &= 0. \end{aligned}$$

- For $0 < \beta \leq 1$: the mixing random variable is $(1-\rho^2)\beta^2 F_0^{2\beta-2} V_T|\sigma_T$. Similarly, we have that

$$\begin{aligned} \hat{\mu}(\sigma_T) &= \ln(F_0 + \theta) + \frac{\rho}{\nu}\beta F_0^{\beta-1}(\sigma_T - \sigma_0), \\ \hat{\beta}(\sigma_T) &= -\frac{1}{2(1-\rho^2)}. \end{aligned}$$

It now remains to derive $\hat{\delta}(\sigma_T)$ and $\hat{\alpha}(\sigma_T)$ by matching the conditional mean and variance of the mixing random variable, i.e.

- For $\beta = 0$

$$\begin{aligned}
\frac{\hat{\delta}(\sigma_T)}{\sqrt{\hat{\alpha}^2(\sigma_T) - \hat{\beta}^2(\sigma_T)}} &= (1-\rho^2)\mathbb{E}(V_T|\sigma_T), \\
\frac{\hat{\delta}(\sigma_T)}{\left(\sqrt{\hat{\alpha}^2(\sigma_T) - \hat{\beta}^2(\sigma_T)}\right)^3} &= (1-\rho^2)^2\mathrm{Var}(V_T|\sigma_T).
\end{aligned}$$

- For $0 < \beta \leq 1$

$$\begin{aligned}
\frac{\hat{\delta}(\sigma_T)}{\sqrt{\hat{\alpha}^2(\sigma_T) - \hat{\beta}^2(\sigma_T)}} &= (1-\rho^2)\beta^2 F_0^{2\beta-2}\mathbb{E}(V_T|\sigma_T), \\
\frac{\hat{\delta}(\sigma_T)}{\left(\sqrt{\hat{\alpha}^2(\sigma_T) - \hat{\beta}^2(\sigma_T)}\right)^3} &= (1-\rho^2)^2\beta^4 F_0^{4\beta-4}\mathrm{Var}(V_T|\sigma_T).
\end{aligned}$$

As there are only two unknowns, solving the above simultaneous equations is a straightforward task.

5.3.3.2 Implementation: two-dimensional integration

Unlike the Normal approximation, we have to perform a two-dimensional integration in order to compute the vanilla call prices using the NIG approximation. Note that as the NIG parameters can be expressed in terms of Z_T, we have that

$$\begin{aligned}
\mathbb{C}_0(K, F_0) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(g(s)-K)^+ f_{NIG}(s;\hat{\alpha}(x),\hat{\beta}(x),\hat{\mu}(x),\hat{\delta}(x))\mathrm{d}s\frac{e^{-\frac{x^2}{2T}}}{\sqrt{2\pi T}}\mathrm{d}x \\
&= \int_{-\infty}^{\infty}\int_{g^{-1}(K)}^{\infty}(g(s)-K) f_{NIG}(s;\hat{\alpha}(x),\hat{\beta}(x),\hat{\mu}(x),\hat{\delta}(x))\mathrm{d}s\frac{e^{-\frac{x^2}{2T}}}{\sqrt{2\pi T}}\mathrm{d}x,
\end{aligned} \tag{5.20}$$

where $g(.)$ (specified in (5.4), (5.5) and (5.6)) is the appropriate transformation and $g^{-1}(.)$ denotes its inverse. Although the above double integral could be a bottleneck in computation and numerically more expensive than the Normal approximation, the implementation scheme is actually quite straightforward. We apply the Simpson's rule, which is found sufficient to give the numerical convergence, to evaluate both the inner and outer integrals. When numerically integrating the outer integral, the upper limit $\overline{z}$ (discussed in the implementation for the Normal approximation) can be taken to be quite comfortably large and we do not have the same problem as the Normal approximation. This is because the growth rate of the inner integral is much slower than the rate of decay of $f_{Z_T}(\cdot)$. Consequently, their product always

tends to zero in the tails of distribution of Z_T. The lower limit $\underline{z}$, on the other hand, has to be chosen with more care. For short maturity, if too low a value of $\underline{z}$ is taken, the NIG parameters can be undefined. This is not really a problem as very small probability mass is assigned to those small values. However, for longer maturity $\underline{z}$ has to be sufficiently small to preserve the probability mass.

Efficiency: One can improve the efficiency of the NIG implementation by the following scheme. Recall that the inner integral of (5.20) has the following form

$$I(x,K) = \int_{g^{-1}(K)}^{\infty} (g(s)-K) f_{NIG}(s;\hat{\alpha}(x),\hat{\beta}(x),\hat{\mu}(x),\hat{\delta}(x))\mathrm{d}s,$$

where f_{NIG} is given by (5.18). For ease of exposition, we write $f_{NIG}(s;\hat{\alpha},\hat{\beta},\hat{\mu},\hat{\delta})$ instead of $f_{NIG}(s;\hat{\alpha}(x),\hat{\beta}(x),\hat{\mu}(x),\hat{\delta}(x))$ but implicitly mean the dependence of the NIG parameters on x. By change of variable, we set

$$y := \hat{\alpha}\hat{\delta}\sqrt{1+\left(\frac{s-\hat{\mu}}{\hat{\delta}}\right)^2}.$$

Hence

$$I(x,K) = \int_{\tilde{l}(x,K)}^{\infty} H(y,K;\hat{\alpha},\hat{\beta},\hat{\mu},\hat{\delta}) K_1(y)\mathrm{d}y,$$

where

$$\begin{aligned}
\tilde{l}(x,K) &= \hat{\alpha}\hat{\delta}\sqrt{1+\left(\frac{g^{-1}(K)-\hat{\mu}}{\hat{\delta}}\right)^2}, \\
H(y,K;\hat{\alpha},\hat{\beta},\hat{\mu},\hat{\delta}) &= \tilde{h}(y,K)\exp\left(\hat{\delta}\sqrt{\hat{\alpha}^2-\hat{\beta}^2}\right)\frac{\exp(\hat{\beta}(s-\hat{\mu}))}{\hat{\delta}^2\sqrt{\left(\frac{y}{\hat{\alpha}\hat{\delta}}\right)^2-1}}, \\
\tilde{h}(y,K) &= g\left(\hat{\mu}+\hat{\delta}\sqrt{\left(\frac{y}{\hat{\alpha}\hat{\delta}}\right)^2-1}\right)-K, \\
s-\hat{\mu} &= \hat{\delta}\sqrt{\left(\frac{y}{\hat{\alpha}\hat{\delta}}\right)^2-1}.
\end{aligned}$$

It can be easily checked that $H(.)$ is a smooth function in y for each fixed set of the NIG parameters and strike K. Therefore, one can approximate $H(.)$ by a piecewise polynomial of the form

$$\begin{aligned}
H(y,K) &= \sum_{n=0}^{m}[a_n(x)-Kb_n(x)]y^n \\
\Rightarrow I(x,K) &\approx \sum_{n=0}^{m}[a_n(x)-Kb_n(x)]\int_{\tilde{l}(x,K)}^{\infty} y^n K_1(y)\mathrm{d}y,
\end{aligned}$$

where the coefficients $\{a_n(x), b_n(x)\}_{n=0}^m$ depend on the NIG parameters. Thus, one can implement the NIG approximation as follows

- For a grid of x values, store the coefficients $\{a_n(x), b_n(x)\}_{n=0}^m$.
- For a grid of l^* values, store the values of the integral $\int_{l^*}^{\infty} y^n K_1(y)\mathrm{d}y$.
- For a given strike K, calculate $I(x, K)$ by using l_i which is the nearest value of l^* to $\tilde{l}(x, K)$, i.e. $l_{i-1} \leq \tilde{l}(x, K) < l_i$

$$\int_{\tilde{l}(x,K)}^{\infty} y^n K_1(y)\mathrm{d}y = \int_{\tilde{l}(x,K)}^{l_i} y^n K_1(y)\mathrm{d}y + \int_{l_i}^{\infty} y^n K_1(y)\mathrm{d}y,$$

  where $\int_{\tilde{l}(x,K)}^{l_i} y^n K_1(y)\mathrm{d}y$ can be evaluated by polynomial interpolation between the grid values l_{i-1} and l_i.

With the above numerical scheme, one can improve the computational efficiency of the two-dimensional integration. Note that this method can also be applied to other European payoff structures.

5.4 Numerical study

In this section, we investigate the quality of the approximations developed in this chapter. It is known that both the Normal and Log-Normal SABR models can be implemented quite well with the SABR formula so we will compare the Normal and NIG approximations with this formula. Similarly for the DD-SABR model, we will test them against the DD-SABR formula.

We take the Monte Carlo solutions (denoted MC for both the Normal and Log-Normal SABR models, and MC-DD for the DD-SABR model) of the SDEs as a natural benchmark to compare all the approximations against. In our numerical study, the initial volatility σ_0 is first chosen to represent the level of the true ATM implied volatility ($\approx \sigma_0 F_0^{\beta-1}$). We force all the ATM implied volatilities produced by the approximations to be the same as the Monte Carlo ATM by adjusting σ_0 and compare errors along the wings as practitioners do in practice.

5.4.1 Normal SABR

We consider the typical parameter values: $\beta = 0, \rho = -0.1, F_0 = 90, \sigma_0 = 9$ for varying maturities T. Since the Normal and NIG approximations work very well for the Normal SABR model, as we shall see in the coming plots, we present our results for the large Volvol cases only and better results are expected to hold for typical market volatility regimes.

The effect on the near the money implied volatility region as maturity increases is illustrated by figures 5.11 and 5.12

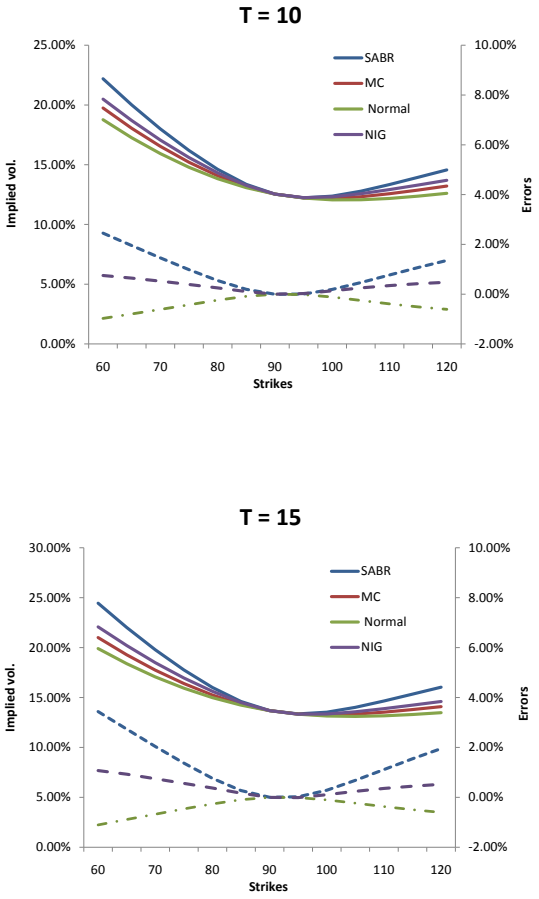


Figure 5.11: Effects of maturity within a high Volvol regime on the Normal and NIG approximations. Other parameters: $\beta = 0, \rho = -0.1, F_0 = 90, \nu = 0.6, \sigma_0 = 9$. The dashed curve of the same colour indicates the errors of the corresponding approximation.

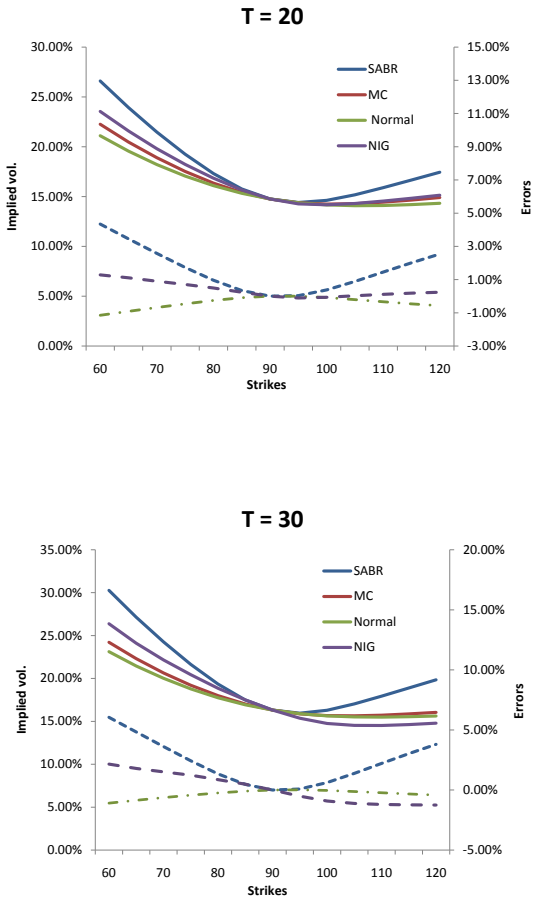


Figure 5.12: Effects of maturity within a high Volvol regime on the Normal and NIG approximations. Other parameters: $\beta = 0, \rho = -0.1, F_0 = 90, \nu = 0.6, \sigma_0 = 9$. The dashed curve of the same colour indicates the errors of the corresponding approximation.

		Strike								
Maturity		30	60	70	80	90	100	110	120	150
10Y	SABR	4.81	2.43	1.44	0.54	0.00	0.19	0.76	1.34	2.64
	Normal	-2.40	-1.00	-0.63	-0.25	0.00	-0.12	-0.39	-0.62	-0.97
	NIG	1.66	0.79	0.54	0.26	0.00	0.15	0.36	0.49	0.67
15Y	SABR	6.97	3.48	2.06	0.77	0.00	0.29	1.13	1.98	3.91
	Normal	-2.87	-1.08	-0.66	-0.26	0.00	-0.10	-0.36	-0.60	-0.99
	NIG	2.85	1.21	0.83	0.40	0.00	0.14	0.41	0.59	0.85
20Y	SABR	8.30	4.36	2.59	0.96	0.00	0.39	1.48	2.57	5.09
	Normal	-3.61	-1.19	-0.71	-0.27	0.00	-0.09	-0.34	-0.58	-1.03
	NIG	3.33	1.37	0.99	0.54	0.00	0.02	0.27	0.43	0.67
30Y	SABR	9.19	6.05	3.62	1.35	0.00	0.62	2.22	3.81	7.47
	Normal	-5.03	-1.20	-0.69	-0.27	0.00	-0.05	-0.26	-0.48	-0.93
	NIG	8.67	2.30	1.62	0.92	0.00	-0.96	-1.15	-1.15	-1.11

Table 5.3: Fitting errors, in percentages, against strike and maturity for $\beta = 0, \nu = 0.6, \rho = -0.1, F_0 = 90, \sigma_0 = 9$ (approximation implied volatility minus MC volatilities).

Comments on the accuracy of approximations: for $\beta = 0$,

- The SABR formula starts losing precision for $T \geq 10$ years while the Normal and NIG approximations still perform quite well and remain relatively close up to 30 years maturity. All the approximations perform worse on the left wing of the implied volatility curves but the errors are still acceptably small for the Normal and NIG approximations (table 5.3). The errors only become substantial when we consider 30 years maturity and low strike (30). Note that in this case, $\nu = 0.6$ represents a highly stress market condition for $T = 10, 15, 20$ and 30 years.

- The Normal approximation does not display enough curvature while the SABR formula shows the opposite. The plots show that it is always a lot closer to the MC solution than the SABR formula on both wings. Furthermore, the errors of the Normal approximation are recorded to be very stable across maturities.

- Similar to the Normal approximation, the NIG approximation works well up to very long maturity even within a high volatility regime, i.e. very high $\nu^2 T \approx 10$. As maturity increases from 20 years to 30 years, the implied volatility curve produced by the NIG approximation becomes progressively

steeper. It is observed in this case that the Normal approximation is a better choice than the NIG approximation.

5.4.2 Log-Normal SABR and DD-SABR

Since the Log-Normal SABR and DD-SABR models yield a lot of similarities in structure, we present their numerical results together and single out the volatility regimes when each individual approximation performs well. We consider the typical parameter values:

- $\beta = 1, F_0 = 90, \rho = -0.5, \sigma_0 = 15\%$.
- $\beta = 0.5, F_0 = 90, \rho = -0.2, \sigma_0 = 130\%$.

Figures 5.13 and 5.14 display the moderate maturity cases where the Normal approximation still performs quite well.

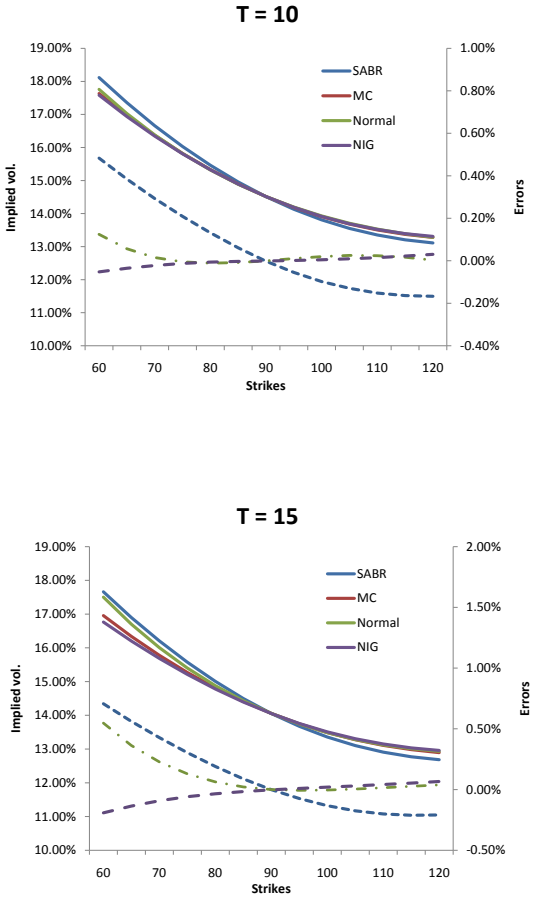


Figure 5.13: Effects of moderate maturity within a low Volvol regime on the Normal and NIG approxmations. Common parameters: $\nu = 0.3, F_0 = 90, \beta = 1, \sigma_0 = 15\%, \rho = -0.5$. The dashed curve of the same colour indicates the errors of the corresponding approximation.

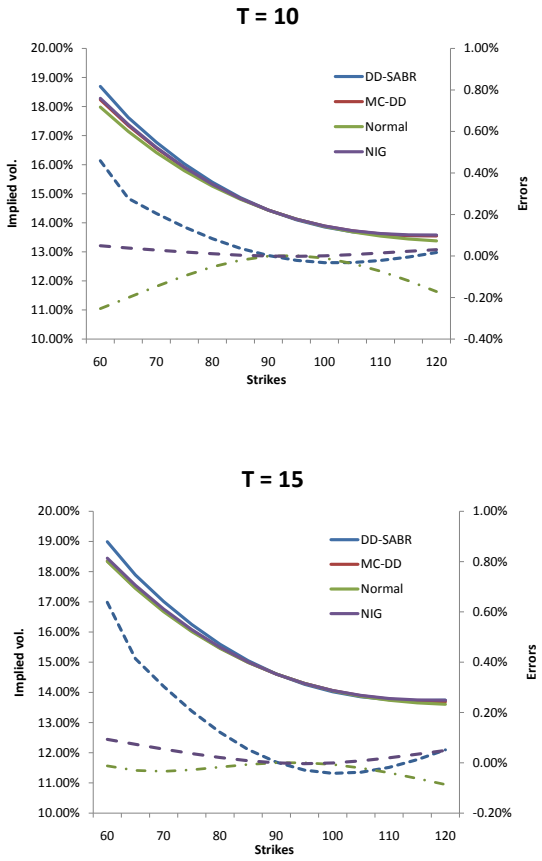


Figure 5.14: Effects of moderate maturity within a low Volvol regime on the Normal and NIG approxmations. Common parameters: $\nu = 0.3, F_0 = 90, \beta = 0.5, \sigma_0 = 130\%, \rho = -0.2$. The dashed curve of the same colour indicates the errors of the corresponding approximation.

When we consider very long maturity or higher Volvol regime with moderate maturity cases ($\nu^2 T > 1.8$), the Normal approximation breaks down due to the reason in remark 9. Although we apply the truncation method mentioned in remark 9, the "Normal" curves are still well above the others. Since the approximation is too far from the true solution, matching the ATM for these cases is a very difficult task. The NIG approximation, on the other hand, still gives very good fits for these cases as illustrated by figures 5.15, 5.16, and 5.17.

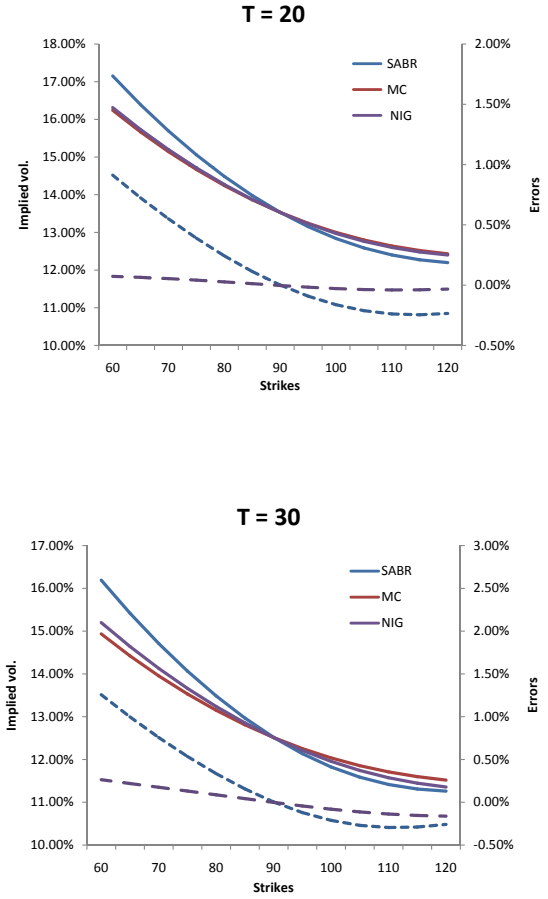


Figure 5.15: Effects of very long maturity within a low Volvol regime on the NIG approximation. Common parameters: $\nu = 0.3, F_0 = 90, \beta = 1, \sigma_0 = 15\%, \rho = -0.5$. The dashed curve of the same colour indicates the errors of the corresponding approximation.

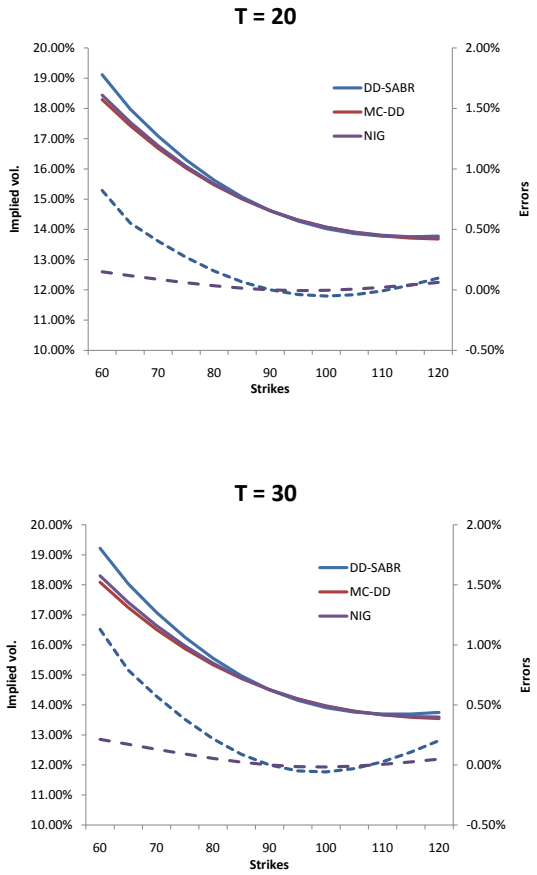


Figure 5.16: Effects of very long maturity within a low Volvol regime on the NIG approximation. Common parameters: $\nu = 0.3, F_0 = 90, \beta = 0.5, \sigma_0 = 130\%, \rho = -0.2$. The dashed curve of the same colour indicates the errors of the corresponding approximation.

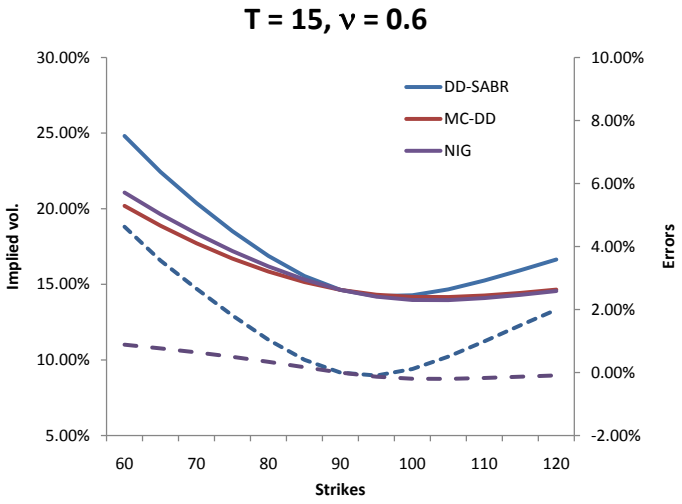


Figure 5.17: Effect of high $\nu^2 T$ (stress volatility regime) on the NIG approximation. Parameters: $\beta = 0.5, \nu = 0.6, F_0 = 90, \sigma_0 = 130\%, \rho = -0.2$. The dashed curve of the same colour indicates the errors of the corresponding approximation.

Comments on the accuracy of approximations: for $0 < \beta \leq 1$

- As β varies from 1 to 0, the Normal and NIG approximations perform better.
- The SABR formula starts breaking down when $T \geq 10$ years or $\nu^2 T \geq 0.9$ as the left and right wings of implied volatility curves are not in line with the MC solutions. On the contrary, the Normal approximation maintains similar shape and therefore fits the MC curves much better than the SABR formula for these cases. It works well up to 15 years and only breaks down for very long maturity or $\nu^2 T > 1.8$.
- The NIG approximation seems to work well up to $\nu^2 T \approx 3.6$ and 5.4 for $\beta = 1$ and 0.5 respectively with the error plots having the lowest magnitude compared with the others. These upper bounds for $\nu^2 T$ are obtained from the following case analysis:
  - Low Volvol ($\nu \approx 0.3$): the NIG approximation performs well up to 30 years maturity for $\beta = 0.5$ and slightly away from the MC solution for $\beta = 1$. Note that in this case, $\nu \approx 0.3$ is the typical market volatility regime for $T \geq 20$ years.
  - High Volvol ($\nu \approx 0.6$): it starts breaking down when $T > 15$ years for $\beta = 0.5$ and $T > 10$ years for $\beta = 1$. The plots show that the errors

are reasonably small with slightly wrong curvature. In this case, $\nu \approx 0.6$ represents the stress volatility regime for moderate maturity.

Remark 10 *In the interest rate area, it is essential to question how the model and the approximations behave for very low rates. As one can expect, the Normal and NIG approximations also assign some positive probability mass to the negative rates region for the DD-SABR model. In table 5.4, we display the mass assigned to negative rates for both approximations and compare with exact (MC-DD) results for the cases considered in Section 5.2.1.2. It is seen from this table that the mass given by the NIG approximation and the MC-DD solution are very close while that given by the Normal approximation is a bit higher. This supports our findings in this section that the NIG approximation is closer in distribution to the MC-DD solution than the Normal approximation.*

T	5 Y	10 Y	20 Y	30 Y
MC-DD	4.03 %	7.67%	19.47%	22.60%
Normal	4.37%	8.22%	22.06%	24.91%
NIG	4.29%	7.49%	18.36%	21.73%

Table 5.4: Probability mass assigned to the negative rates region for the four cases considered in figures 5.7 and 5.8.

5.5 Conclusions

Using an entirely probabilistic framework, we have derived a new approximation for the terminal distribution of the underlying asset. In our method, the main objective is to model the asset's distribution at the maturity date rather than the implied volatilities themselves. This is necessary if we want to extend the approximation to the pricing of more exotic derivatives. The results show that simple approximations which allow for ease of computation are rich enough to capture the model's terminal distribution. The benchmark models we considered in this chapter are the SABR model and the DD-SABR model. In Section 5.2, we find that the CEV-SABR and the DD-SABR model with chosen matching parameters produce very similar implied volatility curves provided that maturity is not too long. Although they are not as close for other cases, we still can work with both models to achieve similar objectives.

In our numerical study, we compare the Normal and NIG approximations with the SABR formula for $\beta = 0, 1$ and the DD-SABR formula for $\beta = 0.5$. When $\beta = 0$, both the Normal and NIG approximations work very well up to 30 years maturity. In the considered stress cases, the Normal approximation is always better

than the SABR formula and remains relatively close to the NIG approximation. Due to its more efficient implementation, the Normal approximation proves to be a very good choice for the Normal SABR model. For $\beta > 0$, the Normal approximation starts losing its precision (slightly away from the true solution) and fails for long maturities (after 20 years) or stress cases (high $\nu^2 T$). However, within its working regimes the Normal approximation still remains better than the SABR formula. We, therefore, conclude that the Normal approximation offers a competitive choice of fitting smiles/skews for short to medium long maturities and normal market condition. The NIG approximation proves to be the best choice here since it outperforms the (DD-)SABR formula and the Normal approximation under all different market scenarios, e.g. normal market condition up to 30 years maturity and stress condition up to 15 years maturity. It also appears to be more theoretically appealing than the Normal approximation although the implementation is slightly trickier to handle. In conclusion, the method addressed in this chapter offers a potentially good approach to other SDE formulations as we shall see in the next chapter.

5.A Appendix: Distribution of F_T under the Log-Normal SABR model

The SDE of the Log-Normal SABR model is

$$\begin{aligned} \mathrm{d}F_t &= \sigma_t F_t \mathrm{d}W_t, \\ \mathrm{d}\sigma_T &= \nu \sigma_t \mathrm{d}Z_t, \\ \mathrm{d}W_t &= \rho \mathrm{d}Z_t + \sqrt{1-\rho^2}\mathrm{d}\hat{W}_t, \end{aligned}$$

where Z and $\hat{W}$ are independent Brownian motions. By Itô's lemma, we have that

$$\begin{aligned} \mathrm{d}\ln F_t &= \frac{1}{F_t}\mathrm{d}F_t - \frac{1}{2}\frac{1}{F_t^2}\mathrm{d}F_t\mathrm{d}F_t \\ \Rightarrow \ln F_T &= \ln F_0 - \frac{1}{2}\int_0^T \sigma_t^2 \mathrm{d}t + \int_0^T \sigma_t \mathrm{d}W_t \\ &= \ln F_0 - \frac{1}{2}\int_0^T \sigma_t^2 \mathrm{d}t + \int_0^T \rho\sigma_t \mathrm{d}Z_t + \sqrt{1-\rho^2}\int_0^T \sigma_t \mathrm{d}\hat{W}_t \\ &= \ln F_0 + \frac{\rho}{\nu}(\sigma_T - \sigma_0) - \frac{1}{2}V_T + \sqrt{1-\rho^2}\int_0^T \sigma_t \mathrm{d}\hat{W}_t, \end{aligned} \tag{5.21}$$

where $V_T := \int_0^T \sigma_t^2 \mathrm{d}t$. Let $M_t := \int_0^t \sigma_u \mathrm{d}\hat{W}_u$ and $\mathcal{F}_T = \sigma(Z_u : 0 \leq u \leq T)$ be the σ-algebra generated by the Brownian motion Z over the time horizon of the option. It is clear that $M_T|\mathcal{F}_T \sim \mathcal{N}(0, V_T)$. By considering the conditional moment

generating function (m.g.f) of M_T, we have that

$$\begin{aligned}\mathbb{E}\left(e^{aM_T}\middle|\mathcal{F}_T\right) &= e^{\frac{1}{2}a^2V_T} \\ &= \mathbb{E}\left(e^{aV_T^{\frac{1}{2}}G}\middle|\mathcal{F}_T\right),\end{aligned}$$

where $G \sim \mathcal{N}(0,1)$ and G is independent of $\mathcal{F}_T$. Consider the m.g.f of $\ln F_T$

$$\begin{aligned}&\mathbb{E}(e^{a\ln F_T}) \\ =\ &\mathbb{E}\left[\mathbb{E}\left\{\exp\left(a\left(\ln F_0 + \frac{\rho}{\nu}(\sigma_T-\sigma_0) - \frac{1}{2}V_T + \sqrt{1-\rho^2}M_T\right)\right)\middle|\mathcal{F}_T\right\}\right] \\ =\ &\mathbb{E}\left[\exp\left(a\left(\ln F_0 + \frac{\rho}{\nu}(\sigma_T-\sigma_0) - \frac{1}{2}V_T\right)\right)\mathbb{E}\left\{\exp\left(a\sqrt{1-\rho^2}M_T\right)\middle|\mathcal{F}_T\right\}\right] \\ =\ &\mathbb{E}\left[\exp\left(a\left(\ln F_0 + \frac{\rho}{\nu}(\sigma_T-\sigma_0) - \frac{1}{2}V_T\right)\right)\mathbb{E}\left\{\exp\left(a\sqrt{1-\rho^2}V_T^{\frac{1}{2}}G\right)\middle|\mathcal{F}_T\right\}\right] \\ =\ &\mathbb{E}\left[\exp\left(a\left(\ln F_0 + \frac{\rho}{\nu}(\sigma_T-\sigma_0) - \frac{1}{2}V_T + \sqrt{1-\rho^2}V_T^{\frac{1}{2}}G\right)\right)\right],\end{aligned}$$

by using the tower and "taking out what is known" properties. Hence

$$\ln F_T \triangleq \ln F_0 + \frac{\rho}{\nu}(\sigma_T-\sigma_0) - \frac{1}{2}V_T + \sqrt{1-\rho^2}V_T^{\frac{1}{2}}G.$$

The same steps follow in both the Normal SABR and DD-SABR models to obtain (5.4) and (5.6) respectively.

5.B Appendix: Proof of Proposition 1 (conditional moments of the realized variance V_T)

To calculate $\mathbb{E}(V_T|\sigma_T)$ and $\mathbb{E}(V_T^2|\sigma_T)$ we use the concept of a Brownian Bridge. By Itô's lemma, we have that:

$$\begin{aligned}\mathrm{d}\ln\sigma_t &= \frac{1}{\sigma_t}\nu\sigma_t\mathrm{d}Z_t - \frac{1}{2\sigma_t^2}\nu^2\sigma_t^2\mathrm{d}t \\ \Rightarrow \ln\sigma_T &= \ln\sigma_0 + \int_0^T \nu\mathrm{d}Z_t - \frac{1}{2}\int_0^T \nu^2\mathrm{d}t \\ \Rightarrow \ln\sigma_T &= \ln\sigma_0 + \nu Z_T - \frac{1}{2}\nu^2 T \\ \iff Z_T &= \frac{\ln\sigma_T - \ln\sigma_0 + \frac{1}{2}\nu^2 T}{\nu}. \end{aligned} \tag{5.22}$$

Hence, if σ_T is known, the value of the end point Z_T is immediate.

Conditional on Z_T, we have a Brownian bridge whose values at time zero

and T are known. Define

$$Z_t|Z_T \triangleq Z_T\frac{t}{T} + (B_t - \frac{t}{T}B_T); \qquad 0 \leq t \leq T \tag{5.23}$$

where B_t is a standard one-dimensional Brownian motion then $Z_t|Z_T$ is a Brownian bridge from 0 to Z_T on $[0,T]$ (Karatzas and Shreve [1991]). It then follows from equation (5.23) that

$$Z_t|\sigma_T = Z_t|Z_T \sim \mathcal{N}\left(\frac{t}{T}Z_T, t - \frac{t^2}{T}\right), \tag{5.24}$$

and it has the following covariance function for $0 \leq t, s \leq T$

$$\mathrm{Cov}(Z_t, Z_s|\sigma_T) = t \wedge s - \frac{ts}{T}. \tag{5.25}$$

5.B.1 First conditional moment of V_T

The conditional expectation of the realized variance can now be written in the following form:

$$\begin{aligned}
\mathbb{E}(V_T|\sigma_T) &= \mathbb{E}\left[\int_0^T \exp\left(2\ln\sigma_0 - \nu^2 t + 2\nu Z_t\right)\mathrm{d}t|\sigma_T\right] \\
&= \int_0^T \mathbb{E}[\exp(2\ln\sigma_0 - \nu^2 t + 2\nu Z_t)|\sigma_T]\mathrm{d}t \\
&= \int_0^T \exp\left(2\ln\sigma_0 - \nu^2 t + \frac{2\nu t}{T}Z_T + 2\nu^2\left(t - \frac{t^2}{T}\right)\right)\mathrm{d}t \\
&= \sigma_0^2 \exp\left(\frac{\left[\frac{\nu\sqrt{T}}{2} + \frac{Z_T}{\sqrt{T}}\right]^2}{2}\right)\sqrt{\frac{\pi T}{2}}\nu^{-1} \\
&\quad \times\left[\Phi\left(\frac{3\nu\sqrt{T}}{2} - \frac{Z_T}{\sqrt{T}}\right) - \Phi\left(-\frac{\nu\sqrt{T}}{2} - \frac{Z_T}{\sqrt{T}}\right)\right],
\end{aligned} \tag{5.26}$$

where $\Phi(\cdot)$ is the cumulative normal distribution function. Plugging back (5.22) to (5.26), we obtain

$$\mathbb{E}(V_T|\sigma_T) = \frac{\sigma_0^2\sqrt{T}}{2\nu}\frac{\left[\Phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}} + \nu\sqrt{T}\right) - \Phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}} - \nu\sqrt{T}\right)\right]}{\phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}} + \nu\sqrt{T}\right)}, \tag{5.27}$$

where $\phi(y) = \frac{e^{\frac{-y^2}{2}}}{\sqrt{2\pi}}$.

5.B.2 Second conditional moment of V_T

We now evaluate the second conditional moment of the realized variance.

$$
\begin{aligned}
& \mathbb{E}(V_T^2|\sigma_T) \\
= \; & \mathbb{E}\left[\left(\int_0^T \sigma_t^2 \mathrm{d}t\right)^2 |\sigma_T\right] \\
= \; & \mathbb{E}\left[2\int_0^T \int_0^t \sigma_t^2\sigma_s^2 \mathrm{d}s\mathrm{d}t|\sigma_T\right] \\
= \; & 2\int_0^T \int_0^t \sigma_0^4 \mathbb{E}\left[\exp(-\nu^2(t+s) + 2\nu(Z_t^{0\to Z_T} + Z_s^{0\to Z_T}))\right] \mathrm{d}s\mathrm{d}t \\
= \; & 2\sigma_0^4 \int_0^T \int_0^t \exp\left(-\nu^2(t+s) + 2\nu\frac{t+s}{T}Z_T + 2\nu^2\left(t - \frac{t^2}{T} + s - \frac{s^2}{T} + 2s - 2\frac{ts}{T}\right)\right) \mathrm{d}s\mathrm{d}t \\
= \; & 2\sigma_0^4 \int_0^T \exp(-4\nu^2 t) \int_0^t \exp\left(-\frac{4\nu^2\frac{(t+s)^2}{T} - 2\frac{2\nu(t+s)}{\sqrt{T}}\left(\frac{5\nu\sqrt{T}}{2} + \frac{Z_T}{\sqrt{T}}\right)}{2}\right) \mathrm{d}s\mathrm{d}t. \qquad (5.28)
\end{aligned}
$$

By completing the square and change of variable $u = \frac{2\nu(t+s)}{\sqrt{T}} - \left(\frac{5\nu\sqrt{T}}{2} + \frac{Z_T}{\sqrt{T}}\right)$ in the inner integral of (5.28), we have that

$$
\begin{aligned}
& \mathbb{E}(V_T^2|\sigma_T) \\
= \; & 2\sigma_0^4 \exp\left(\frac{\left(\frac{5\nu\sqrt{T}}{2} + \frac{Z_T}{\sqrt{T}}\right)^2}{2}\right) \sqrt{\frac{\pi T}{2}}\nu^{-1} \int_0^T \exp(-4\nu^2 t) \int_{\frac{2\nu t - Z_T}{\sqrt{T}} - \frac{5\nu}{2}\sqrt{T}}^{\frac{4\nu t - Z_T}{\sqrt{T}} - \frac{5\nu}{2}\sqrt{T}} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \mathrm{d}u\mathrm{d}t \\
= \; & 2\sigma_0^4 \exp\left(\frac{\left(\frac{5\nu\sqrt{T}}{2} + \frac{Z_T}{\sqrt{T}}\right)^2}{2}\right) \sqrt{\frac{\pi T}{2}}\nu^{-1} \\
& \times \int_0^T \exp(-4\nu^2 t) \underbrace{\left[\Phi\left(\frac{4\nu t - Z_T}{\sqrt{T}} - \frac{5\nu}{2}\sqrt{T}\right) - \Phi\left(\frac{2\nu t - Z_T}{\sqrt{T}} - \frac{5\nu}{2}\sqrt{T}\right)\right]}_{:=\Phi(g_1(t)) - \Phi(g_2(t))} \mathrm{d}t.
\end{aligned}
$$

The above integral can be evaluated by integration by parts

$$
\begin{aligned}
&\int_0^T \exp(-4\nu^2 t)\Phi(g_1(t))\mathrm{d}t \\
= \;& -\frac{\exp(-4\nu^2 t)}{4\nu^2}\Phi(g_1(t))\Big|_0^T + \int_0^T \frac{\exp(-4\nu^2 t)}{4\nu^2} g_1^{'}(t)\phi(g_1(t))\mathrm{d}t \\
= \;& \left[-\frac{\exp(-4\nu^2 T)}{4\nu^2}\Phi\left(\frac{-Z_T}{\sqrt{T}}+\frac{3\nu}{2}\sqrt{T}\right)+\frac{1}{4\nu^2}\Phi\left(\frac{-Z_T}{\sqrt{T}}-\frac{5\nu}{2}\sqrt{T}\right)\right] \\
& +\frac{1}{\nu\sqrt{T}}\int_0^T \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\left(\frac{4\nu t-Z_T}{\sqrt{T}}-\frac{5\nu}{2}\sqrt{T}\right)^2+8\nu^2 t}{2}\right)\mathrm{d}t \\
= \;& \left[-\frac{\exp(-4\nu^2 T)}{4\nu^2}\Phi\left(\frac{-Z_T}{\sqrt{T}}+\frac{3\nu}{2}\sqrt{T}\right)+\frac{1}{4\nu^2}\Phi\left(\frac{-Z_T}{\sqrt{T}}-\frac{5\nu}{2}\sqrt{T}\right)\right] \\
& +\frac{1}{\nu\sqrt{T}}\exp\left(\frac{\left(\frac{3\nu\sqrt{T}}{2}+\frac{Z_T}{\sqrt{T}}\right)^2}{2}-\frac{\left(\frac{5\nu\sqrt{T}}{2}+\frac{Z_T}{\sqrt{T}}\right)^2}{2}\right) \\
& \times\int_0^T \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\left(\frac{4\nu t-Z_T}{\sqrt{T}}-\frac{3\nu}{2}\sqrt{T}\right)^2}{2}\right)\mathrm{d}t. \qquad (5.29)
\end{aligned}
$$

By change of variable $v=\frac{4\nu t-Z_T}{\sqrt{T}}-\frac{3\nu}{2}\sqrt{T}$ in (5.29), we find that

$$
\begin{aligned}
&\int_0^T \exp(-4\nu^2 t)\Phi(g_1(t))\mathrm{d}t \\
= \;& \left[-\frac{\exp(-4\nu^2 T)}{4\nu^2}\Phi\left(\frac{-Z_T}{\sqrt{T}}+\frac{3\nu}{2}\sqrt{T}\right)+\frac{1}{4\nu^2}\Phi\left(\frac{-Z_T}{\sqrt{T}}-\frac{5\nu}{2}\sqrt{T}\right)\right] \\
& +\frac{1}{4\nu^2}\exp\left(\frac{\left(\frac{3\nu\sqrt{T}}{2}+\frac{Z_T}{\sqrt{T}}\right)^2}{2}-\frac{\left(\frac{5\nu\sqrt{T}}{2}+\frac{Z_T}{\sqrt{T}}\right)^2}{2}\right) \\
& \times\left[\Phi\left(\frac{-Z_T}{\sqrt{T}}+\frac{5\nu}{2}\sqrt{T}\right)-\Phi\left(\frac{-Z_T}{\sqrt{T}}-\frac{3\nu}{2}\sqrt{T}\right)\right].
\end{aligned}
$$

Similarly,

$$\begin{aligned}
&\int_0^T \exp(-4\nu^2 t)\Phi(g_2(t))\mathrm{d}t \\
= &\left[-\frac{\exp(-4\nu^2 T)}{4\nu^2}\Phi\left(\frac{-Z_T}{\sqrt{T}}-\frac{\nu}{2}\sqrt{T}\right)+\frac{1}{4\nu^2}\Phi\left(\frac{-Z_T}{\sqrt{T}}-\frac{5\nu}{2}\sqrt{T}\right)\right] \\
&+\frac{1}{4\nu^2}\exp\left(\frac{\left(\frac{\nu\sqrt{T}}{2}+\frac{Z_T}{\sqrt{T}}\right)^2}{2}-\frac{\left(\frac{5\nu\sqrt{T}}{2}+\frac{Z_T}{\sqrt{T}}\right)^2}{2}\right) \\
&\times\left[\Phi\left(\frac{-Z_T}{\sqrt{T}}+\frac{3\nu}{2}\sqrt{T}\right)-\Phi\left(\frac{-Z_T}{\sqrt{T}}-\frac{\nu}{2}\sqrt{T}\right)\right].
\end{aligned}$$

Putting all the pieces together we obtain

$$\begin{aligned}
\mathbb{E}(V_T^2|\sigma_T) = &-\frac{\sigma_0^4\sqrt{T}}{4\nu^3}\left(1+e^{2\ln(\sigma_T/\sigma_0)}\right)\frac{\left[\Phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}}+\nu\sqrt{T}\right)-\Phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}}-\nu\sqrt{T}\right)\right]}{\phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}}+\nu\sqrt{T}\right)} \\
&+\frac{\sigma_0^4\sqrt{T}}{4\nu^3}\frac{\left[\Phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}}+2\nu\sqrt{T}\right)-\Phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}}-2\nu\sqrt{T}\right)\right]}{\phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}}+2\nu\sqrt{T}\right)}. \qquad (5.30)
\end{aligned}$$

5.C Appendix: Proof of proposition 2 (conditional mean and variance of s_T)

We prove the Log-Normal SABR case only as similar calculations apply to other models. The **conditional mean** of s_T is

$$\begin{aligned}
\mu(\sigma_T) &= \mathbb{E}(s_T|\sigma_T) \\
&= \ln F_0+\frac{\rho}{\nu}(\sigma_T-\sigma_0)-\frac{1}{2}\mathbb{E}(V_T|\sigma_T)+\sqrt{1-\rho^2}\mathbb{E}(V_T^{\frac{1}{2}}G|\sigma_T). \qquad (5.31)
\end{aligned}$$

Recall that $\mathcal{F}_T=\sigma(Z_u:0\le u\le T)$. It follows that

$$\begin{aligned}
\mathbb{E}(V_T^{\frac{1}{2}}G|\sigma_T) &= \mathbb{E}(\mathbb{E}(V_T^{\frac{1}{2}}G|\mathcal{F}_T)|\sigma_T) \\
&= \mathbb{E}(V_T^{\frac{1}{2}}\mathbb{E}(G|\mathcal{F}_T)\mid\sigma_T) \\
&= \mathbb{E}(V_T^{\frac{1}{2}}\mathbb{E}(G)\mid\sigma_T) \\
&= 0.
\end{aligned}$$

Hence

$$\mu(\sigma_T)=\ln F_0+\frac{\rho}{\nu}(\sigma_T-\sigma_0)-\frac{1}{2}\mathbb{E}(V_T|\sigma_T). \qquad (5.32)$$

The **conditional variance** of s_T:

$$\begin{aligned}\eta^2(\sigma_T) &= \mathrm{Var}(s_T|\sigma_T) \\ &= \frac{1}{4}\mathrm{Var}(V_T|\sigma_T) - \sqrt{1-\rho^2}\mathrm{Cov}(V_T, V_T^{\frac{1}{2}}G|\sigma_T) + (1-\rho^2)\mathrm{Var}(V_T^{\frac{1}{2}}G|\sigma_T).\end{aligned} \tag{5.33}$$

Similarly, the covariance term in (5.33) can be expressed as

$$\begin{aligned}\mathrm{Cov}(V_T, V_T^{\frac{1}{2}}G|\sigma_T) &= \mathbb{E}(V_T^{\frac{3}{2}}G|\sigma_T) - \mathbb{E}(V_T|\sigma_T)\mathbb{E}(V_T^{\frac{1}{2}}G|\sigma_T) \\ &= \mathbb{E}(V_T^{\frac{3}{2}}G|\sigma_T) \\ &= 0,\end{aligned}$$

and the last term in (5.33) is

$$\begin{aligned}\mathrm{Var}(V_T^{\frac{1}{2}}G|\sigma_T) &= \mathbb{E}(V_T G^2|\sigma_T) - [\mathbb{E}(V_T^{\frac{1}{2}}G|\sigma_T)]^2 \\ &= \mathbb{E}(V_T G^2|\sigma_T) - 0 \\ &= \mathbb{E}(V_T|\sigma_T),\end{aligned}$$

again using the tower property and noting that $\mathbb{E}(G^2) = 1$. Hence

$$\eta^2(\sigma_T) = \frac{1}{4}\mathbb{E}(V_T^2|\sigma_T) - \frac{1}{4}[\mathbb{E}(V_T|\sigma_T)]^2 + (1-\rho^2)\mathbb{E}(V_T|\sigma_T). \tag{5.34}$$

Given the formulae from the previous appendix, the results follow immediately.

5.D Appendix: DD-SABR equivalent Black implied volatility

In this appendix, we derive an equivalent Black implied volatility formula for the DD-SABR model using the techniques developed in Hagan et al. [2002] but with a few modifications from later literature, e.g. Hagan et al. [2005], Obloj [2008]. We start with a more general form of the SABR model:

$$\begin{aligned}\mathrm{d}F_t &= \hat{\sigma}_t C(F_t)\mathrm{d}W_t, \\ \mathrm{d}\hat{\sigma}_t &= \nu\hat{\sigma}_t\mathrm{d}Z_t, \\ \mathrm{d}Z_t\mathrm{d}W_t &= \rho\mathrm{d}t,\end{aligned}$$

where the function $C(u)$ is is assumed to be positive, smooth and integrable around 0:

$$\int_0^x \frac{\mathrm{d}u}{C(u)} < \infty, \qquad x > 0.$$

Equation (B.65) in Appendix B of Hagan et al. [2002] yields the equivalent Black implied volatility for the above model:

$$\begin{aligned} & \sigma_B(K,F_0) \\ = \; & \frac{\hat{\sigma}_0 \ln F_0/K}{\int_K^{F_0} \frac{\mathrm{d}u}{C(u)}} \left(\frac{z}{x(z)}\right) \times \\ & \left\{1 + \left[\frac{2\gamma_2 - \gamma_1^2 + \frac{1}{F_{av}^2}}{24}\hat{\sigma}_0^2 C^2(F_{av}) + \frac{1}{4}\rho\nu\hat{\sigma}_0\gamma_1 C(F_{av}) + \frac{2-3\rho^2}{24}\nu^2\right] T + \ldots\right\}. \end{aligned} \tag{5.35}$$

Here

$$F_{av} = \sqrt{F_0 K}, \tag{5.36}$$

$$\gamma_1 = \frac{C^{'}(F_{av})}{C(F_{av})}, \tag{5.37}$$

$$\gamma_2 = \frac{C^{''}(F_{av})}{C(F_{av})}, \tag{5.38}$$

and

$$\begin{aligned} z &= \frac{\nu}{\hat{\sigma}_0}\frac{F_0 - K}{C(F_{av})}, \\ x(z) &= \ln\left\{\frac{\sqrt{1-2\rho z + z^2} + z - \rho}{1-\rho}\right\}, \end{aligned}$$

where $\frac{z}{x(z)}$ basically represents the main effect of the stochastic volatility. The fraction $\frac{z}{x(z)}$ is taken to be 1 for the ATM case in the limit sense for proper ATM calibration. There are quite a few criticisms of this function z, e.g. Obloj [2008]. We use a more general form of z as proposed in Hagan et al. [2005].

$$z = \frac{\nu}{\hat{\sigma}_0}\int_K^{F_0} \frac{\mathrm{d}u}{C(u)}. \tag{5.39}$$

Note that for this choice of z, implied volatilities obtained by this approximation coincide with Berestycki et al. [2004] and Obloj [2008] for the CEV-SABR model. We now turn our attention to the DD-SABR model as introduced in the main chapter. The model is the special case:

$$\begin{aligned} C(F_t) &= F_t + \theta, \\ \theta &= F_0\frac{1-\beta}{\beta}, \\ \hat{\sigma}_0 &= \sigma_0 \beta F_0^{\beta-1}. \end{aligned}$$

Making this substitution in (5.37), (5.38) and (5.39) we obtain:

$$\begin{aligned}\gamma_1 &= \frac{1}{\sqrt{F_0K}+\theta},\\ \gamma_2 &= 0,\\ z &= \frac{\nu}{\hat\sigma_0}\ln\frac{F_0+\theta}{K+\theta}.\end{aligned}$$

Substituting further in (5.35), we get:

$$\begin{aligned}\sigma_B(K,F_0) &= \frac{\hat\sigma_0 \ln F_0/K}{\ln(F_0+\theta)/(K+\theta)}\left(\frac{z}{x(z)}\right)\{1+M\cdot T+\dots\},\\ M &= \frac{-1/(\sqrt{F_0K}+\theta)^2+1/(F_0K)}{24}\hat\sigma_0^2(\sqrt{F_0K}+\theta)^2\\ &\quad +\frac{1}{4}\rho\nu\hat\sigma_0\frac{1}{\sqrt{F_0K}+\theta}(\sqrt{F_0K}+\theta)+\frac{2-3\rho^2}{24}\nu^2\\ &= \frac{-1+(\sqrt{F_0K}+\theta)^2/(F_0K)}{24}\hat\sigma_0^2+\frac{1}{4}\rho\nu\hat\sigma_0+\frac{2-3\rho^2}{24}\nu^2\\ &= \frac{2\theta/\sqrt{F_0K}+\theta^2/(F_0K)}{24}\hat\sigma_0^2+\frac{1}{4}\rho\nu\hat\sigma_0+\frac{2-3\rho^2}{24}\nu^2.\end{aligned}$$

We can simplify this formula by expanding[4]

$$\begin{aligned}&(F_0+\theta)-(K+\theta)\\ =\;&\sqrt{(F_0+\theta)(K+\theta)}\ln\frac{F_0+\theta}{K+\theta}\left\{1+\frac{1}{24}\ln^2\frac{F_0+\theta}{K+\theta}+\frac{1}{1920}\ln^4\frac{F_0+\theta}{K+\theta}+\dots\right\},\\ &F_0-K\\ =\;&\sqrt{F_0K}\ln\frac{F_0}{K}\left\{1+\frac{1}{24}\ln^2\frac{F_0}{K}+\frac{1}{1920}\ln^4\frac{F_0}{K}+\dots\right\}.\end{aligned}$$

Hence, the implied volatility formula now reads

$$\begin{aligned}\sigma_B(K,F_0) &= \hat\sigma_0\frac{\sqrt{(F_0+\theta)(K+\theta)}}{\sqrt{F_0K}}\left(\frac{1+\frac{1}{24}\ln^2\frac{F_0+\theta}{K+\theta}+\frac{1}{1920}\ln^4\frac{F_0+\theta}{K+\theta}+\dots}{1+\frac{1}{24}\ln^2\frac{F_0}{K}+\frac{1}{1920}\ln^4\frac{F_0}{K}+\dots}\right)\left(\frac{z}{x(z)}\right)\\ &\quad \{1+\left[\frac{2\theta/\sqrt{F_0K}+\theta^2/(F_0K)}{24}\hat\sigma_0^2+\frac{1}{4}\rho\nu\hat\sigma_0+\frac{2-3\rho^2}{24}\nu^2\right]T+\dots\}.\end{aligned}$$

For the special case of ATM options, we first take the limit

$$\lim_{K\to F_0}\frac{\ln F_0/K}{\ln(F_0+\theta)/(K+\theta)}=\frac{F_0+\theta}{F_0},\tag{5.40}$$

[4]We use Hagan's technique with the assumption that the strike K is not so far away from F_0

and hence the formula reduces to

$$
\begin{aligned}
& \sigma_B(F_0, F_0) \\
= \;& \hat{\sigma}_0 \frac{F_0+\theta}{F_0} \left\{ 1 + \left[\frac{2\theta/F_0 + \theta^2/F_0^2}{24} \hat{\sigma}_0^2 + \frac{1}{4}\rho\nu\hat{\sigma}_0 + \frac{2-3\rho^2}{24}\nu^2 \right] T + \ldots \right\} \\
= \;& \sigma_0 F_0^{\beta-1} \left\{ 1 + \left[\frac{2\frac{1-\beta}{\beta} + \frac{(1-\beta)^2}{\beta^2}}{24} \sigma_0^2 \beta^2 F_0^{2\beta-2} + \frac{1}{4}\rho\nu\sigma_0\beta F_0^{\beta-1} + \frac{2-3\rho^2}{24}\nu^2 \right] T + \ldots \right\} \\
= \;& \sigma_0 F_0^{\beta-1} \left\{ 1 + \left[\frac{1-\beta^2}{24} \sigma_0^2 F_0^{2\beta-2} + \frac{1}{4}\rho\nu\sigma_0\beta F_0^{\beta-1} + \frac{2-3\rho^2}{24}\nu^2 \right] T + \ldots \right\}.
\end{aligned}
$$

Chapter 6

On the approximation of the SABR with mean reversion model: a probabilistic approach

In this chapter, we study the SABR with mean reversion model (SABR-MR). We first compare the SABR model with the SABR-MR model in terms of forward volatility to point out the fundamental difference in the models' dynamics. We then derive an efficient probabilistic approximation for the SABR-MR model to price European options. Similar to the method derived in Chapter 5, we focus on capturing the terminal distribution of the underlying asset (conditional on the terminal volatility) to arrive at the implied volatilities of the corresponding European options for all strikes and maturities. As in the SABR case, the method allows us to work with a wide range of parameters which cover long dated options and different market conditions.

6.1 Introduction

Stochastic volatility modelling is a demanding area and the modelling requirements vary from market to market. As noted in Chapter 5, the SABR model that was originally proposed in Hagan et al. [2002] is amongst the most popular stochastic volatility models in practice. It is widely used to model the forward price of a stock or the forward LIBOR/swap rates in the fixed income market. In the interest rate context, relevant market (caplet/swaption) prices or market smiles will inform us about the marginal distribution of the underlying rate at its start date (under some martingale measure) and the SABR model can calibrate well to this information. The downside of the model, as argued by a number of authors, is the fact that we need different SABR models (different parameters sets) to recover market marginals

of different rates (at different start dates). This leads to an incoherent model that isolates the rates from each other and hence one cannot use the SABR dynamics to model a common driver for many rates. A comprehensive discussion on this topic can be found in Rebonato et al. [2009]. While it might not be a problem for the pricing of European swaptions which relies on the distribution of only one rate at its start date under a certain martingale measure, systematic problems may be encountered as we want to apply the model to price other derivatives. Examples of such products include payoffs that depend on more than one rate or one rate but at more than one future date.

Having listed a few drawbacks of the SABR model, we want to switch focus to a model extension that could prove to be practically useful: the SABR with mean reversion model (SABR-MR). To be specific, the usual geometric Brownian motion for the volatility process in the SABR model is replaced by the product of the exponential of an Ornstein Uhlenbeck process and an extra deterministic function of time. Incorporating mean reversion into the volatility process has been previously adopted by a number of authors. For example, Fouque et al. [2000a], Masoliver and Perello [2006] and Perello et al. [2008] studied the equity market and investigated why the mean reverting behaviour of the volatility process is needed to explain the data (S&P 500 and Dow Jones Industrial Average indices). In case of the interest rate market, the works by Kaisajuntti and Kennedy [2011], Piterbarg [2005] and the most recent paper by Antonov and Spector [2012] also agree that the mean reverting volatility setup is needed especially for large time horizons. A particular form of the SABR-MR model was studied in Kaisajuntti and Kennedy [2011] within the context of swaption data. The authors found that mean reversion will be essential if one wishes to provide a low dimensional model for the level of rates. In terms of the option pricing problem, it was investigated and concluded in this paper that the SABR and SABR-MR models can price European options and calibrate to market smiles of all expiries equally well. The model introduced in Piterbarg [2005] is similar to the SABR-MR model intuitively but the SDE formulation is a bit different. Instead of using the exponential Ornstein Uhlenbeck process, the author assumed that the stochastic variance follows the mean reverting square-root dynamics. The main objective is then to apply this proposed SDE formulation to develop a full term structure model.

While agreeing that both the SABR and SABR-MR models can price European options equally well, we want to extend the analysis further and give some thought as to why the SABR-MR model might be more suitable for other products. In Section 6.3, we examine the difference in forward volatility (from future time t to expiry T) between the two models, which determines the forward smiles. Derivatives that are exposed to forward volatility include from quite simple contracts such as

forward start options (swaptions), forward swaption straddles, to more complicated structures such as cliquet options.

The second objective of the chapter is to derive an efficient approximation for the SABR-MR model to price European options. We adopt a similar method to the one derived in Chapter 5 for the SABR model but with some further adjustments. As in Chapter 5, this method focuses on the marginal distribution of the underlying at option's maturity to arrive at the required implied volatilities. Once we fit an appropriate approximation to the underlying's marginal distribution, implied volatilities can be immediately recovered by inverting the option prices. In Chapter 5, we found that this approach was significantly better than the benchmark approximation proposed in Hagan et al. [2002]. An advantage for the probabilistic approximation in Chapter 5 is that it can cope well with very long maturity (up to 30 years) or high volatility of volatility. This is a different feature from other current approaches in the literature (applied to the SABR model) that rely on the assumption of very small total volatility and usually degrades for longer than 10 years maturity, e.g. Hagan et al. [2002], Hagan et al. [2005], Wu [2010], Henry-Labordere [2005] and Paulot [2009]. The works that we mentioned earlier, i.e. Fouque et al. [2000a] and Perello et al. [2008], also developed some approximations for pricing European options within the exponential Ornstein Uhlenbeck stochastic volatility model. However, the proposed methods are only valid for a range of typical parameters in a certain market (very short dated options). Later in the chapter, we will compare our approximation with the true Monte Carlo solution to assess its quality.

Here we briefly review the main features of the probabilistic approach used in Chapter 5 that we intend to apply to the SABR-MR model. The first important remark is that we keep the general correlation structure of the model to retain the flexibility of capturing the skews. We then build up our approximation by conditioning the underlying's distribution on the terminal volatility and approximating this distribution. The resulting approximate conditional distribution (with approximately correct mean and variance) has to be theoretically appealing (close to the true distribution) but simple enough to allow for computational efficiency. As done in Chapter 5, we propose the Normal Inverse Gaussian distribution for such purposes.

Secondly, a challenge for this approach is the constant elasticity of variance (CEV) structure of the SABR-MR model which admits no explicit solution. As in Chapter 5, we study the simpler displaced diffusion (DD) model where our previously mentioned method can be applied. There, the CEV and DD versions of the SABR model have been compared numerically and the results show that the two models continue to be qualitatively similar. In this chapter, we continue to adopt this mapping with the aim of transferring the intuition from the CEV to DD version of

the SABR-MR model.

The chapter is organized as follows. Section 6.3 compares the SABR and SABR-MR models in terms of forward volatility through some numerical examples. We develop our approximation in Section 6.4. In Section 6.5, we numerically investigate the quality of our approximation and compare against Monte Carlo simulations as the benchmark. Section 6.6 concludes the chapter.

6.2 An alternative stochastic volatility model for modelling smiles

This section starts by reviewing the SABR model.

6.2.1 The SABR and SABR with mean reversion models (SABR-MR)

Recall that under the SABR model, the dynamics of the underlying asset[1] is given by

$$\begin{aligned} \mathrm{d}y_t &= \sigma_t y_t^{\beta} \mathrm{d}W_t \qquad \beta \in [0,1], \\ \mathrm{d}\sigma_t &= \nu \sigma_t \mathrm{d}Z_t \qquad \nu > 0, \end{aligned} \tag{6.1}$$

where W_t and Z_t are correlated Brownian motions such that $\mathrm{d}W_t \mathrm{d}Z_t = \rho \mathrm{d}t$ for all $t \leq T$ with $\rho \in [-1, 1]$. The model assumes that the underlying process is already a (local) martingale under some equivalent martingale measure. Each parameter in the SABR model has a specific role in determining the shapes of the skews and smiles. See Hagan et al. [2002] and Chapter 5 for these roles.

Although the SABR model is a powerful tool for vanilla swaption pricing, the biggest limitation is that it can only price one single vanilla at a time. To be specific, for each swap rate of different start date one will have to feed a different set of SABR parameters into the model in order to reproduce the correct market smile (see Rebonato et al. [2009]). For this reason, the dynamics of the SABR model might not be an appropriate choice to describe the market implied distributions of the swap rates at different time points in the future.

In order to better understand the point above, we assess the SABR model from a different viewpoint. Today's market smile of a swap rate that expires in T years tells us everything about today's marginal distribution of $(y_T | y_0 = y, \sigma_0 = \sigma)$ and the SABR model can get this right. While this task can be achieved by more than one stochastic volatility model, it does not guarantee that we will also get right

[1] Note that we switch the notation for the underlying from F as in Chapter 5 to y. This is because we work more with interest rate data so it seems more suitable to use the chosen notation.

the forward conditional distribution of $(y_T|y_t, \sigma_t)$ for $t \in (0, T)$ which can be crucial for other forward volatility sensitive products. We tackle this issue in Section 6.3.

We now discuss an extension of the SABR model. One choice is to incorporate mean reversion into the volatility process. Instead of the Log-Normal specification as in the SABR model, we assume that volatility is the exponential of an Ornstein Uhlenbeck (mean reverting) process U. Before going into details of the main model, we first review some existing works that consider similar formulations.

Adding mean reversion to volatility is not new in the area. A number of authors have already considered this formulation but with different objectives. The first model that is worth mentioning is the exponential Ornstein Uhlenbeck model proposed by Fouque et al. [2000a]. In this model, only $\beta = 1$ is considered in the SDE for the underlying and the volatility process as we mentioned above is of the form $\exp(U_t)$ where

$$\mathrm{d}U_t = -\kappa U_t \mathrm{d}t + \nu \mathrm{d}Z_t \qquad \nu, \kappa > 0. \tag{6.2}$$

They then applied the model in the equity market with a strong focus on the S&P 500 index. In such a market, they observed that volatility is fast mean-reverting when looked at over the time scale of a European derivative contract (many months but not years). Using an asymptotic method, they derived an implied volatility formula to price European options on the S&P 500 index but only for a certain range of parameters that fit this market. Extending the analysis in Fouque et al. [2000a] but for the interest rate derivatives, in Cotton et al. [2004] the authors considered simple short rate models such as CIR and Vasicek with the presence of fast mean-reverting stochastic volatility and applied to price bond options. Two other papers that are also similar in scope and setup are the works by Masoliver and Perello [2006] and Perello et al. [2008]. In Masoliver and Perello [2006], by observing annualized volatility from intensive historical data (Dow Jones Industrial Average index from 1900 to 2004) the authors spotted the long term memory property of the volatility, i.e. it reverts to the stationary distribution after a certain length of time. This research gives further strong empirical evidence that the exponential Ornstein Uhlenbeck model is desirable in such a market. In Perello et al. [2008], they again looked at the option pricing problem under this model but focused on a wider range of parameters than that considered in Fouque et al. [2000a]. Furthermore, the estimation of parameters comes directly from the historical data observation of long term volatility, rate of mean reversion and correlation with the log-returns, etc. Since the approximation derived in Perello et al. [2008] is only valid for very short dated options, we find it irrelevant for our main application which includes pricing long dated European options so we will not discuss it further in this chapter.

We now look at our main model of interest, the SABR with mean reversion model which will be referred to as the SABR-MR model. The form of the SABR-MR

model we choose to study is as in Kaisajuntti and Kennedy [2011]

$$\begin{aligned} \mathrm{d}y_t &= \sigma_t y_t^{\beta} \mathrm{d}W_t \qquad \beta \in [0,1], \\ \sigma_t &= I_t \exp(U_t), \\ \mathrm{d}U_t &= -\kappa U_t \mathrm{d}t + \nu \mathrm{d}Z_t \qquad \nu, \kappa > 0, \qquad U_0 = 0, \\ \mathrm{d}W_t \mathrm{d}Z_t &= \rho \mathrm{d}t \qquad \rho \in [-1,1], \end{aligned} \tag{6.3}$$

where $I_t = a\exp\{-c(T-t)\}$ is a deterministic function of time with $a > 0$ being the initial volatility and $c > 0$ being the rate of exponential decay. Similar to the initial volatility σ_0 in the SABR model, the function I_t also has the role of controlling the ATM level of the smile. Furthermore, this choice of function is a lot more flexible as it also controls the expected level of the volatility process through time. Note that we have the freedom to choose other forms of the function I_t for the SABR-MR model.

Note that U is an Ornstein Uhlenbeck process with κ being the mean reversion parameter. A large κ will pull U fast towards its mean reversion level which is set to be zero in this model. This implies that the volatility process $\sigma_t = I_t \exp(U_t)$ of the SABR-MR model will mean revert around I_t for $t \in [0,T]$. A reasonably large level of κ ensures that the volatility of the underlying will not be too volatile and grow too fast with expiry. When $\kappa \to 0$ we have $\mathrm{Var}(\ln \sigma_T) = \nu^2 T$ which is the same as in the SABR model.

In the interest rate context, this specific form of model was first introduced in Kaisajuntti and Kennedy [2011] with the original motivation coming from a numerical investigation of 10-years tenor swap rates of different start dates (2 to 30 years) in order to identify a model for the level of rates. As discussed in Kaisajuntti and Kennedy [2011], ν has the role of controlling the overall curvature of different smiles across expiries while κ controls the dampening over time. In other words, a reasonable κ will flatten out the smile curve as maturity increases. The roles of other parameters are found to be similar to those in the SABR model.

6.2.1.1 The DD-SABR-MR model

As noted in Chapter 5, the power law type function $\phi(y) = y^{\beta}$ as specified in (6.1) and (6.3) can enable a very flexible modelling of volatility skew. However, the CEV structure does not admit an explicit solution and it is not very straightforward numerically to implement. Similar to what we did in Chapter 5, we consider a modification of (6.3), the displaced diffusion SABR-MR model (DD-SABR-MR). In fact, this is also the main model of interest in Kaisajuntti and Kennedy [2011]. In

this model, the dynamics of the underlying asset is given as

$$\begin{aligned} \mathrm{d}y_t &= \hat{\sigma}_t(y_t+\theta)\mathrm{d}W_t, \\ \sigma_t &= I_t \exp(U_t), \\ \mathrm{d}U_t &= -\kappa U_t \mathrm{d}t + \nu \mathrm{d}Z_t \qquad U_0 = 0, \\ \mathrm{d}W_t \mathrm{d}Z_t &= \rho \mathrm{d}t, \end{aligned} \tag{6.4}$$

for some displaced diffusion coefficient $\theta \in \mathbb{R}$. We apply the mapping as introduced in Marris [1999] to match up the CEV structure in (6.3) when $\beta \in (0,1)$

$$\begin{aligned} \hat{\sigma}_t &= \sigma_t \beta y_0^{\beta-1}, \\ \theta &= y_0 \frac{1-\beta}{\beta}. \end{aligned}$$

It is well known from Marris [1999] and Svoboda-Greenwood [2009] that in the deterministic volatility case, the dynamics in (6.3) and (6.4) with the above mapping are very similar and implied volatilities produced by the two models are almost identical across a wide range of strikes. In Chapter 5, it is found that this is still the case qualitatively for the SABR model and its displaced diffusion version (DD-SABR). We, therefore, assume the close link between the CEV-SABR-MR model (specified in (6.3)) and the DD-SABR-MR model (specified in (6.4)) to hold qualitatively. For ease of notation, we will continue to refer to the parameter β for the DD-SABR-MR model and implicitly use the above mapping. The other two special cases will be referred to as the Normal SABR-MR model ($\beta = 0$) and the Log-Normal SABR-MR model ($\beta = 1$). Later in some parts of the chapter, we employ the terminologies SABR and SABR-MR but we will mean the families of the Normal, Log-Normal, and DD versions of the SABR and SABR-MR models respectively unless otherwise specified.

6.3 Forward volatility

Up to now, we have described the SABR-MR model in terms of its capability of pricing European options. We recall that what matters for this problem is the distribution of $(y_T|y_0 = y, \sigma_0 = \sigma)$. The fact that the SABR and SABR-MR models can give reasonably similar prices for the same European option implies that they generate similar distribution of $(y_T|y_0 = y, \sigma_0 = \sigma)$ which in turn determines today's market smile. However, it does not tell us anything about whether the two models will give the same (conditional) forward smiles, i.e. the distribution of $(y_T|y_t, \sigma_t)$, $t \in (0,T)$. This conditional distribution, of course, depends on the volatility paths $\{\sigma_u\}, u \in [t,T]$. In the literature, these volatility paths are often referred to as the

forward volatility. For the two models, the two different autonomous SDEs for the volatility process potentially imply very different dynamics of the forward volatility. In what follows, we compare them numerically to pull out the differences that can have some level of practical significance.

6.3.1 Products

We now consider a toy example of forward start options which involves the study of forward volatility. Let us first define this product.

A forward start option in general is a contract in which the holder receives, at future time T_1 (at no additional cost), an option with expiry date $T_2 > T_1$. Here we consider the simplest example of a forward start call option, i.e. the payoff at time T_2 is $(y_{T_2} - Ky_{T_1})^+$ which depends on the realizations of the underlying at two future dates T_1 and T_2. The strike ratio K here could be: < 1 (ITM), $= 1$ (ATM), or > 1 (OTM). Prices of this product for various K are determined by the forward smile. One can also define other payoffs such as: $(Ky_{T_1} - y_{T_2})^+$ for put, $|y_{T_2} - y_{T_1}|$ for straddle.

Pricing methodology: in order to price a forward start option, we use Monte Carlo simulation to generate different realizations of the underlying at T_1 and T_2, i.e. $y_{T_1}(\omega_i)$ and $y_{T_2}(\omega_i)$ where each ω_i indicates a particular outcome. The numeraire-rebased price of this product at time T_0 can be written as

$$\begin{aligned} \mathbb{C}_0(T_1, T_2, y_0, \sigma_0) &= \mathbb{E}[(y_{T_2} - Ky_{T_1})^+] \\ &\approx \frac{1}{N}\sum_{i=1}^{N}(y_{T_2}(\omega_i) - Ky_{T_1}(\omega_i))^+, \end{aligned} \tag{6.5}$$

where N is the number of paths used for Monte Carlo simulation.

Calibration: for accurate valuation, one has to calibrate the model properly to the right market information. To be specific, we need to consider the followings for our application

- The distribution of $(y_{T_2}|y_0 = y, \sigma_0 = \sigma)$: this is implied by today's T_2-market smile.
- The distribution of $(y_{T_1}|y_0 = y, \sigma_0 = \sigma)$: this is implied by today's T_1-market smile.
- Having specified both $(y_{T_2}|y_0 = y, \sigma_0 = \sigma)$ and $(y_{T_1}|y_0 = y, \sigma_0 = \sigma)$, we are then left with $(y_{T_2}|y_{T_1}, \sigma_{T_1})$ which is fully determined by the forward volatility. Note that this information is forced to model users as long as the first two requirements are fulfilled.

6.3.2 Some numerical examples

Having mentioned the calibration requirements for forward start option pricing, we now proceed to the set up of the model. Suppose we are given two market smiles for European options expiring at T_1 and T_2 for calibration, we wish to investigate if the price of a forward start option changes due to the change of model (SABR to SABR-MR). Recall that the main feature that differentiates the SABR-MR model from the SABR model is the mean reverting behaviour of the volatility process. This behaviour is controlled by the parameter κ. Hence, it is sensible to view prices as a function of κ and assess how it changes with respect to this parameter. Of course, as κ approaches zero, the model dynamics becomes closer to the SABR dynamics. This can be seen by looking at the SDE (6.2) for U_t of the SABR-MR model. When $\kappa = 0$, the SDE for $\exp(U_t)$ is the same as that in the SABR model plus a deterministic drift term.

To fit the model to the T_1 and T_2-smiles using different choices of κ, we alter the volatility process $\{\sigma_t\}_{t\in[0,T_2]}$ of the original SABR-MR model as follows

$$\begin{aligned} \mathrm{d}U_t &= -\kappa U_t \mathrm{d}t + \nu(t)\mathrm{d}Z_t, \qquad U_0 = 0, \\ \sigma_t &= \alpha(t)\exp(U_t), \end{aligned}$$

where $\nu(t)$ and $\alpha(t)$ are both piecewise-constant

$$\begin{aligned} \nu(t) &= \nu_1 \qquad t \in [0, T_1), \\ &= \nu_{12} \qquad t \in [0, T_2], \\ \alpha(t) &= \alpha_1 \qquad t \in [0, T_1), \\ &= \alpha_{12} \qquad t \in [0, T_2]. \end{aligned}$$

All constants here must be positive. In this choice, the function $\alpha(t)$ is used instead of the original deterministic function I_t for simplification. With two different constants α_1 and α_{12} for two time periods, we ensure that the ATM levels of T_1 and T_2-smiles are properly matched. The volatility of volatility function $\nu(t)$ is used to guarantee that the smiles can also be fitted properly along the wings. It is clear that large values of ν_1 and ν_{12} are accompanied with large κ for accurate calibration.

Our numerical experiment will be carried out in a number of steps. For a chosen β, we generate two arbitrary symmetric (i.e. $\rho = 0$ for simplification) smiles for European options with: the underlying swap rate $y_0 = 5\%$, expiries $T_1 = 5\text{Y}$, $T_2 = 10\text{Y}$, and assume they are given for calibration. We then fit the above model to these two smiles for various values of κ and calculate the forward start option prices using (6.5). Three different values of strike ratios K are used in this example

- $K = 1$ (ATM).
- $K = 0.75$ (ITM).
- $K = 1.25$ (OTM).

The prices and calibrated parameters are plotted for different values of κ in the following figures. The result for the case $\beta = 1$ is observed to be qualitatively similar so we will not display it here.

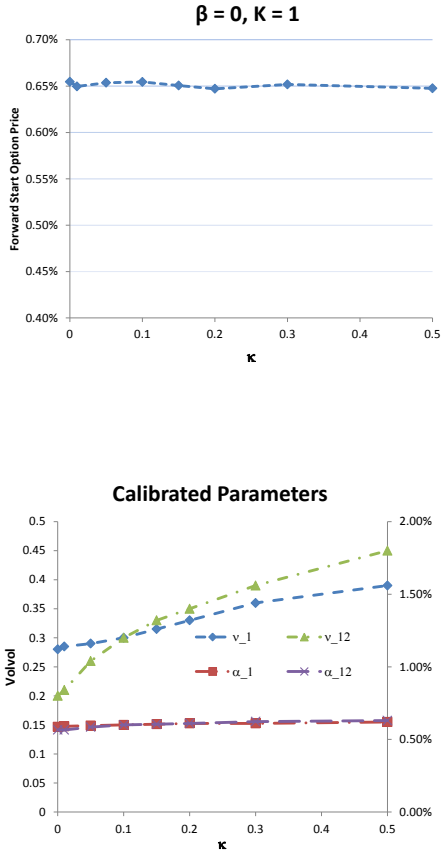
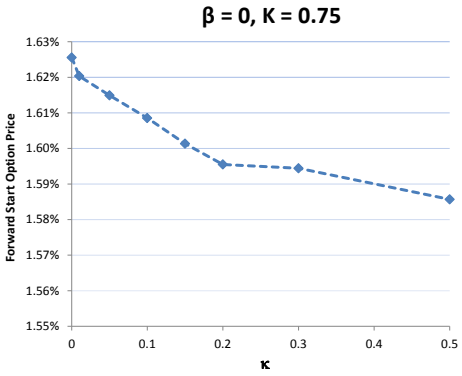


Figure 6.1: Prices of forward start options for various κ: $\beta = 0, \rho = 0$, $y_0 = 5\%, T_2 = 10\text{Y}$, $T_1 = 5\text{Y}$, $K = 1$ and calibrated parameters.



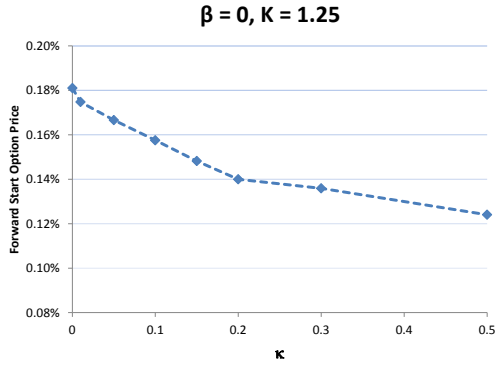


Figure 6.2: Prices of forward start options for various κ: $\beta = 0, \rho = 0$, $y_0 = 5\%, T_2 = 10\text{Y}$, $T_1 = 5\text{Y}$, $K = 0.75$ and $K = 1.25$.

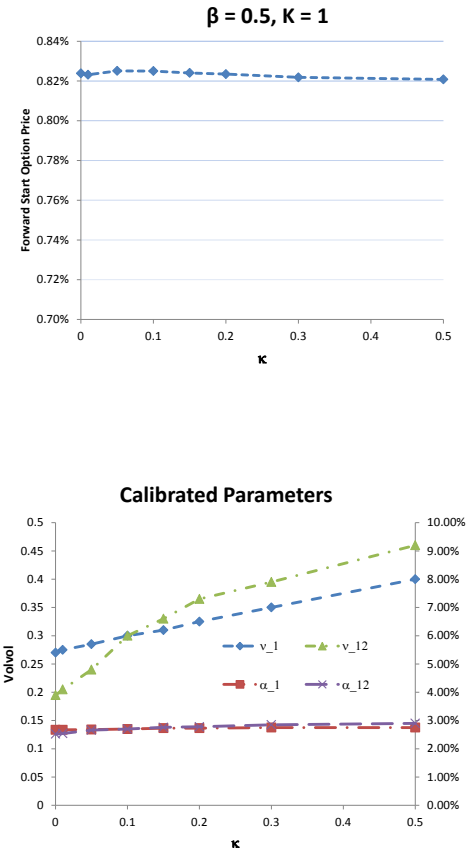
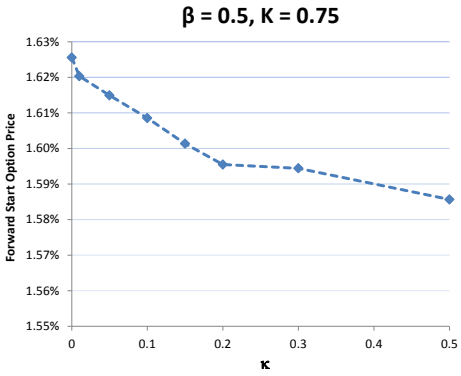


Figure 6.3: Prices of forward start options for various κ: $\beta = 0.5, \rho = 0$, $y_0 = 5\%, T_2 = 10\text{Y}$, $T_1 = 5\text{Y}$, $K = 1$ and calibrated parameters.



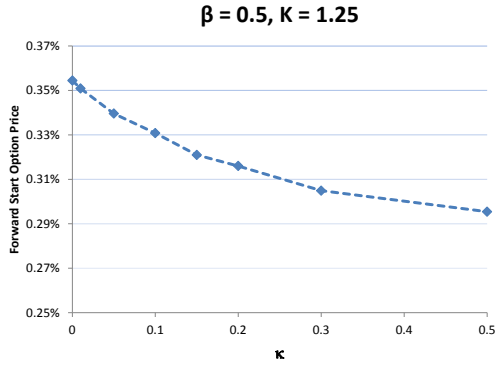


Figure 6.4: Prices of forward start options for various κ: $\beta = 0.5, \rho = 0$, $y_0 = 5\%, T_2 = 10\text{Y}$, $T_1 = 5\text{Y}$, $K = 0.75$ and $K = 1.25$.

From the results, one can immediately spot that although prices are not so variant with κ for the ATM case, the price curves for the ITM and OTM cases are clearly monotonically decreasing in κ. This is an interesting result since it clearly differentiates the SABR dynamics ($\kappa = 0$) from the SABR-MR dynamics ($\kappa > 0$).

To see how forward volatility might have some influence on the price of a forward start option, we look at the distributions of the underlying rate at both dates T_1 and T_2. For the simplest case ($\beta = 0, \rho = 0$) that we considered, we can write down the exact distributions of y_{T_1} and y_{T_2} as follows (see Appendices 6.A and 6.B for derivation)

$$\begin{aligned} y_{T_1} &\triangleq y_0 + \sqrt{V_{T_1}} G_1 \quad , \qquad V_{T_1} = \int_0^{T_1} \sigma_t^2 \mathrm{d}t, \\ y_{T_2} &\triangleq y_{T_1} + \sqrt{V_{T_2}^1} G_{12} \quad , \qquad V_{T_2}^1 = \int_{T_1}^{T_2} \sigma_t^2 \mathrm{d}t, \\ & G_1 \perp G_{12} \quad , \quad G_1, G_{12} \sim \mathcal{N}(0,1), \end{aligned} \tag{6.6}$$

and G_1 and G_{12} are also independent of $\{\sigma_t\}_{t\in[0,T_2]}$. We term $V^1_{T_2}$ by the "forward" realized variance to indicate that the integral of the variance process starts from future date T_1 to T_2. Since the conditional distribution of $(y_{T_1}|y_0 = y, \sigma_0 = \sigma)$ is similar for all levels of κ due to the calibration to T_1-market smile, heuristically the difference in price tends to come from the difference in the conditional distribution of $(y_{T_2}|y_{T_1}, \sigma_{T_1})$. From (6.6), it is intuitively clear that the forward realized variance $V^1_{T_2}$ is one of the main ingredients to control the distribution of $(y_{T_2}|y_{T_1}, \sigma_{T_1})$ which in turn affects the forward start option prices. We plot the first and second moments of the forward realized variance below to illustrate this intuition.

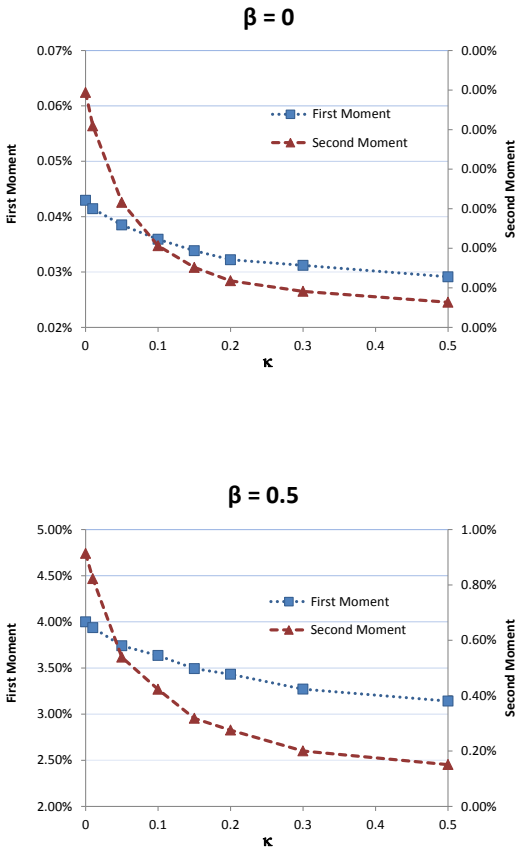


Figure 6.5: First and second moments of the forward realized variance $V^1_{T_2} = \int_{T_1}^{T_2} \sigma_t^2 \mathrm{d}t$ for the considered cases.

6.4 A probabilistic approximation for the SABR-MR model

In this section, we develop an approximation for the SABR-MR model described in Section 6.2.1 following a similar approach in Chapter 5. Although the main objective

is the pricing of European swaptions, the proposed approach can easily be applied for equity and other markets as mentioned in the literature review in Section 6.2.1.

6.4.1 Approximating the terminal distribution

We recall at this point the SDE formulations (6.3) for the Normal and Log-Normal versions ($\beta = 0, 1$), and (6.4) for the DD version ($\beta \in (0, 1)$) of the SABR-MR model. By the fundamental pricing formula and tower property, today's numeraire-rebased price of a vanilla call option struck at some strike K is given by

$$\begin{aligned} \mathbb{C}_0(K, y_0) &= \mathbb{E}\left[(y_T - K)^+\right] \\ &= \mathbb{E}\left[\mathbb{E}\{(y_T - K)^+ | U_T\}\right]. \end{aligned} \tag{6.7}$$

Assuming that we have in mind some distribution for $y_T|U_T$: the conditional distribution of y_T given U_T, then the conditional expectation above can be evaluated as a double integral since the law of U_T is known (see (6.2) and Appendix 6.C). To keep the notation simple and transparent, we introduce the process s that represents the level of assets and function $g(.)$ to transform it back to the underlying process y, i.e. $y_t = g(s_t)$. As the first stepping stone, we will write down the exact solutions in distribution to the SDE for our reference models (see Appendix 6.A for details).

- **Normal SABR-MR**:

$$\begin{aligned} s_T &\triangleq y_0 + \frac{\rho}{\nu}(F(T, U_T) - F(0, U_0)) + \rho H_T + \sqrt{1-\rho^2} V_T^{\frac{1}{2}} G, \\ g(s) &= s. \end{aligned} \tag{6.8}$$

- **Log-Normal SABR-MR**:

$$\begin{aligned} s_T &\triangleq \ln y_0 + \frac{\rho}{\nu}(F(T, U_T) - F(0, U_0)) + \rho H_T - \frac{1}{2} V_T + \sqrt{1-\rho^2} V_T^{\frac{1}{2}} G, \\ g(s) &= e^s. \end{aligned} \tag{6.9}$$

- **DD-SABR-MR**:

$$\begin{aligned} s_T &\triangleq \ln(y_0 + \theta) + \beta y_0^{\beta-1} \frac{\rho}{\nu}(F(T, U_T) - F(0, U_0)) + \rho\beta y_0^{\beta-1} H_T - \frac{1}{2}\beta^2 y_0^{2\beta-2} V_T \\ &\quad + \sqrt{1-\rho^2}\beta y_0^{\beta-1} V_T^{\frac{1}{2}} G, \\ \theta &= y_0 \frac{1-\beta}{\beta}, \\ g(s) &= e^s - \theta, \end{aligned} \tag{6.10}$$

where $F(t, U_t) = I_t \exp(U_t)$ is the anti-derivative of $I_t \exp(U_t)$ with respect to U_t and

$$V_T := \int_0^T {I_t}^2 \exp(2U_t)\mathrm{d}t, \tag{6.11}$$

$$H_T := -\left(\frac{\nu}{2} + \frac{c}{\nu}\right)\int_0^T I_t \exp(U_t)\mathrm{d}t + \frac{\kappa}{\nu}\int_0^T I_t \exp(U_t)U_t\mathrm{d}t, \tag{6.12}$$

and G is a standard normal random variable that is independent of U, i.e. it is also independent of V_T and H_T. Here, we refer to the random variable V_T as the realized variance.

We note that this stage is a bit different from the SABR case in Chapter 5. Conditional on $U_T = u^*$, the above solutions in distribution will involve the random variables H_T, V_T and an independent Normal variable G. It is seen that H_T has a direct connection with the correlation parameter ρ and hence a certain control over the slope of the implied volatility curve. At an intuitive level, an initial guess is that H_T plays a similar role to the second term involving $F(T, U_T)$ in (6.8), (6.9) and (6.10), and thus does not display too much variability once conditioned on $U_T = u^*$. Therefore, we first approximate $H_T|U_T = u^*$ by its conditional mean $\mathbb{E}(H_T|U_T = u^*)$ and treat it as a deterministic term. Our approximation will now become of the form

$$s_T|U_T = u^* \approx \tilde{s}_T|U_T = u^*,$$

where $\tilde{s}_T|U_T$ now involves the conditional mean $\mathbb{E}(H_T|U_T = u^*)$ instead of $H_T|U_T = u^*$. Note that when $\rho = 0$, $s_T|U_T = u^*$ and $\tilde{s}_T|U_T = u^*$ coincide. We now proceed to the second step of the approximation where the conditional mean and variance of $\tilde{s}_T|U_T = u^*$ need to be evaluated. We then aim to approximate the conditional distribution of $\tilde{s}_T|U_T = u^*$ by replacing it with some suitable random variable with the same conditional mean and variance. In each case, H_T and the realized variance V_T play a central role in our calculation and analysis so we will treat its moments separately in the following proposition.

Proposition 3 *Assume that the dynamics of the volatility is given by* $I_t \exp(U_t)$ *where*

$$\begin{aligned} I_t &= a\exp\{-c(T-t)\} \qquad a, c > 0, \\ \mathrm{d}U_t &= -\kappa U_t \mathrm{d}t + \nu \mathrm{d}Z_t \qquad \nu, \kappa > 0, \end{aligned}$$

where $U_0 = 0$ *and* $t \in [0, T]$. *Given the covariance function of* U

$$\begin{aligned} R(t,s) &= Cov(U_t, U_s) = \frac{\nu^2}{2\kappa}[\exp\{-\kappa|t-s|\} - \exp\{-\kappa(t+s)\}], \\ R(t,t) &= R(t), \end{aligned}$$

the first ***conditional moments of*** H_T ***and the realized variance*** V_T *have the following integral expressions*

$$\mathbb{E}(V_T|U_T = u^*) = \int_0^T (I_t)^2 \exp\left(2u^* \frac{R(t,T)}{R(T)} + 2R(t) - 2\frac{R^2(t,T)}{R(T)}\right) \mathrm{d}t, \quad (6.13)$$

$$\mathbb{E}(H_T|U_T = u^*) = -\left(\frac{\nu}{2} + \frac{c}{\nu}\right) H_1(u^*) + \frac{\kappa}{\nu} H_2(u^*), \quad (6.14)$$

where

$$\begin{aligned} H_1(u^*) &= \int_0^T I_t \exp\left(u^* \frac{R(t,T)}{R(T)} + \frac{R(t) - \frac{R^2(t,T)}{R(T)}}{2}\right) \mathrm{d}t, \\ H_2(u^*) &= \int_0^T I_t \left(u^* \frac{R(t,T)}{R(T)} + R(t) - \frac{R^2(t,T)}{R(T)}\right) \exp\left(u^* \frac{R(t,T)}{R(T)} + \frac{R(t) - \frac{R^2(t,T)}{R(T)}}{2}\right) \mathrm{d}t. \end{aligned}$$

The second ***conditional moment of the realized variance*** V_T *can be expressed as the following double integral*

$$\begin{aligned} & \mathbb{E}(V_T^2 \mid U_T = u^*) \\ = \; & 2\int_0^T (I_t)^2 \int_0^t (I_s)^2 \exp\left(2u^* \left(\frac{R(t,T)}{R(T)} + \frac{R(s,T)}{R(T)}\right) + 2(R(t) + R(s))\right) \\ \times \; & \exp\left(-2\frac{R^2(t,T) + R^2(s,T)}{R(T)} + 4R(t,s) - 4\frac{R(t,T)R(s,T)}{R(T)}\right) \mathrm{d}s\mathrm{d}t. \end{aligned}$$

Proof: see Appendix 6.C.

Remark 11 *When* $\kappa, c = 0$, *we obtain analytical expressions for the first and second conditional moments of* V_T *as the volatility dynamics coincides with the SABR volatility dynamics.*

6.4.2 Normal Inverse Gaussian approximation

In this section, we consider the Normal Inverse Gaussian distribution (NIG) for the approximation of $\tilde{s}_T|U_T$ (defined in the previous section) rather than $s_T|U_T$ as in the SABR case. This choice of distribution was proposed in Chapter 5 together with the Normal distribution. There, the numerical investigation showed that the NIG approximation was found to be superior compared with the other. We have tested

both approximations for the SABR-MR model and since the NIG approximation continues to be much better, we will consider it only.

Under the NIG approximation, we assume

$$\tilde{s}_T|U_T = u^* \sim \mathcal{NIG}(\hat{\alpha}, \hat{\beta}, \hat{\mu}, \hat{\delta}),$$

where the parameters are to be chosen. The NIG density function $f_{NIG}(s; \hat{\alpha}, \hat{\beta}, \hat{\mu}, \hat{\delta})$ is the same as (5.18) in Chapter 5.

6.4.2.1 Matching parameters

Similar to Chapter 5, we now describe an efficient way to match the NIG parameters. We again use the fact that a NIG random variable X can be expressed as the Normal variance-mean mixture form:

$$X = \hat{\mu} + \hat{\beta}Y + \sqrt{Y}G, \tag{6.15}$$

where the mixing random variable Y follows an Inverse Gaussian (IG) distribution and G is a standard Normal random variable that is independent of Y. It is clear from (6.8) to (6.10) that conditioned on $U_T = u^*$, $\tilde{s}_T$ as an approximation of s_T will have a similar form as (6.15). We will now express the NIG parameters in terms of u^*.

Recall that we replace $\tilde{s}_T|U_T = u^*$ with a NIG random variable, i.e.

$$\tilde{s}_T|U_T = u^* \sim \mathcal{NIG}(\hat{\alpha}(u^*), \hat{\beta}(u^*), \hat{\mu}(u^*), \hat{\delta}(u^*)).$$

Matching parameters for the NIG approximation in the SABR-MR model is the same as for the SABR model, as shown in Chapter 5, provided that we work with the approximate random variable $\tilde{s}_T|U_T = u^*$ instead of $s_T|U_T = u^*$.

- For $\beta = 0$: the mixing random variable is $(1-\rho^2)V_T|U_T = u^*$. We first match the location and asymmetry parameters

$$\begin{aligned}\hat{\mu}(u^*) &= y_0 + \frac{\rho}{\nu}(F(T,u^*) - F(0,0)) + \rho\mathbb{E}(H_T \mid U_T = u^*),\\ \hat{\beta}(u^*) &= 0.\end{aligned}$$

- For $0 < \beta \leq 1$: the mixing random variable is $(1-\rho^2)\beta^2 y_0^{2\beta-2}V_T|U_T = u^*$. Similarly, we have that

$$\begin{aligned}\hat{\mu}(u^*) &= \ln(y_0+\theta) + \frac{\rho}{\nu}\beta y_0^{\beta-1}(F(T,u^*) - F(0,0)) + \rho\beta y_0^{\beta-1}\mathbb{E}(H_T \mid U_T = u^*),\\ \hat{\beta}(u^*) &= -\frac{1}{2(1-\rho^2)}.\end{aligned}$$

It now remains to derive $\hat{\delta}(u^*)$ and $\hat{\alpha}(u^*)$ by matching the conditional mean and variance of the mixing random variable, i.e.

- For $\beta = 0$

$$\begin{aligned}
\frac{\hat{\delta}(u^*)}{\sqrt{\hat{\alpha}^2(u^*) - \hat{\beta}^2(u^*)}} &= (1-\rho^2)\mathbb{E}(V_T|U_T = u^*), \\
\frac{\hat{\delta}(u^*)}{\left(\sqrt{\hat{\alpha}^2(u^*) - \hat{\beta}^2(u^*)}\right)^3} &= (1-\rho^2)^2\mathrm{Var}(V_T|U_T = u^*).
\end{aligned}$$

- For $0 < \beta \leq 1$

$$\begin{aligned}
\frac{\hat{\delta}(u^*)}{\sqrt{\hat{\alpha}^2(u^*) - \hat{\beta}^2(u^*)}} &= (1-\rho^2)\beta^2 y_0^{2\beta-2}\mathbb{E}(V_T|U_T = u^*), \\
\frac{\hat{\delta}(u^*)}{\left(\sqrt{\hat{\alpha}^2(u^*) - \hat{\beta}^2(u^*)}\right)^3} &= (1-\rho^2)^2\beta^4 y_0^{4\beta-4}\mathrm{Var}(V_T|U_T = u^*).
\end{aligned}$$

As there are only two unknowns, solving the above simultaneous equations is a straightforward task.

6.4.2.2 Implementation

In order to implement the NIG approximation, we have to perform a two-dimensional integration in order to compute the vanilla call option price.

$$\begin{aligned}
\mathbb{C}_0(K, y_0) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(g(s)-K)^+ f_{NIG}(s;\hat{\alpha}(u),\hat{\beta}(u),\hat{\mu}(u),\hat{\delta}(u))\mathrm{d}s\frac{e^{-\frac{u^2}{2R(T)}}}{\sqrt{2\pi R(T)}}\mathrm{d}u \\
&= \int_{-\infty}^{\infty}\int_{g^{-1}(K)}^{\infty}(g(s)-K) f_{NIG}(s;\hat{\alpha}(u),\hat{\beta}(u),\hat{\mu}(u),\hat{\delta}(u))\mathrm{d}s\frac{e^{-\frac{u^2}{2R(T)}}}{\sqrt{2\pi R(T)}}\mathrm{d}u,
\end{aligned} \tag{6.16}$$

where $f_{NIG}(.)$ is the density function of corresponding NIG random variable; $g(.)$ is the appropriate transformation for the chosen β and $g^{-1}(.)$ denotes its inverse. Although the above double integral could be a bottleneck in computation and numerically expensive, the implementation scheme is actually quite straightforward. We apply the Simpson's rule, which is found sufficient to give the numerical convergence, to evaluate both the inner and outer integrals.

Efficiency: One can improve the efficiency of the NIG implementation by the interpolating polynomials scheme mentioned in Chapter 5. However, the integral

forms of $\mathbb{E}(V_T|U_T = u^*), \mathbb{E}(V_T^2|U_T = u^*)$ and $\mathbb{E}(H_T|U_T = u^*)$ in Proposition 3 make the NIG approximation for the SABR-MR a bit slower to implement than that for the SABR model as in Chapter 5.

6.5 Numerical study

In this section, we test the quality of the NIG approximation under different volatility regimes. We take the Monte Carlo solutions (MC) of the SDEs as a natural benchmark to compare the NIG approximation against. In our numerical study, the parameter a is first chosen to represent the average level of the true ATM implied volatilities. We force all the ATM-implied volatilities produced by the NIG approximation to be the same as the MC-ATM by adjusting the parameter a. We then compare the approximation errors along the wings by subtracting MC implied volatility from the approximate implied volatility. Swaption prices are calculated for strikes -200, -100, -50, -25, 0, 25, 50, 100, 200 basis points (bp) away from the ATM strike (y_0). Note that we test the approximation for typical swaption data but the same conclusions apply to other markets too.

6.5.1 Normal SABR-MR ($\beta = 0$)

In this section, we investigate the quality of the NIG approximation for the Normal SABR-MR model. We consider the following typical parameter values: $\beta = 0, y_0 = 0.05, a = 0.0075$ for varying T. In the results shown below, we vary one parameter at a time to assess the effect of such parameter on the approximation. For each parameter, we display the smiles produced by the MC solution and the NIG's errors (approximation minus MC) for $T = 2, 5, 10, 20$, and 30 years (Y).

The effect of ν: we first test the parameters which control the curvature of the smile. Recall that ν controls the overall curvature of the smiles while κ represents the dampening effect through time. The effect of ν on the NIG approximation is displayed in figures 6.6 and 6.7. Two values of $\nu = 0.25, 0.4$ are tested while other parameters are held fixed: $\kappa = 0.05, c = 0.1, \rho = -0.1$. One can see that as ν varies from 0.4 to 0.25, the magnitude of the errors across expiries decreases dramatically. For the worst case when $T = 30$Y and strike is 0.03 (-200 bp strike offset), the size of the error drops by about five and a half times ($0.90\% \to 0.16\%$).

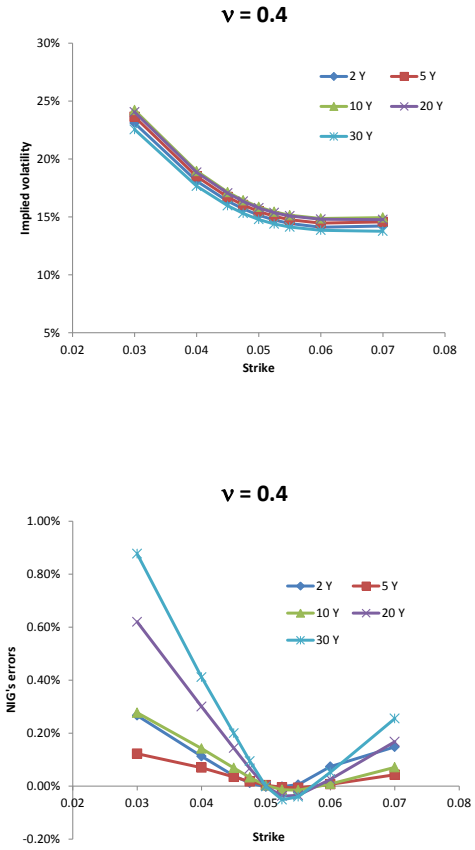


Figure 6.6: ν-effect on the NIG approximation: $\nu = 0.4$ (MC solution and approximation errors). Other parameters: $\beta = 0, y_0 = 0.05, \kappa = 0.05, c = 0.1, \rho = -0.1$.

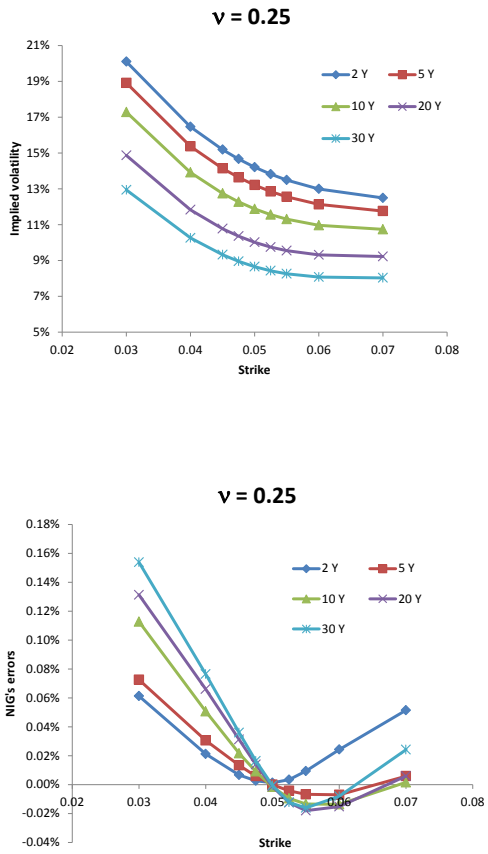


Figure 6.7: ν-effect on the NIG approximation: $\nu = 0.25$ (MC solution and approximation errors). Other parameters: $\beta = 0, y_0 = 0.05, \kappa = 0.05, c = 0.1, \rho = -0.1$.

The effect of κ: we now fix $\nu = 0.4$ and other parameters $c = 0.1, \rho = -0.1$ but increases κ from 0.05 to 0.15. As κ controls the curvature through time, one expects that as κ gets larger smiles get flat faster. We observe that smiles with $\nu = 0.4, \kappa = 0.15$ look quite similar to those with $\nu = 0.25, \kappa = 0.05$ and indeed the errors are also of similar size although they seem to be a bit larger in the right wing for all expiries.

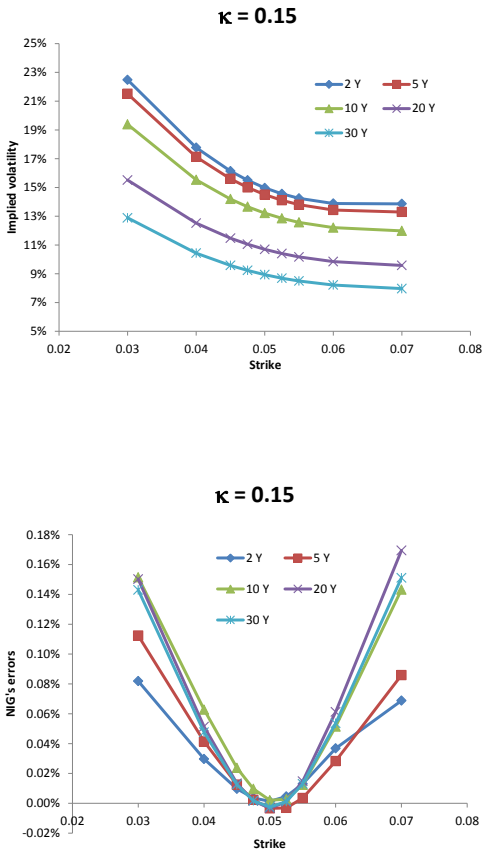


Figure 6.8: κ-effect on the NIG approximation (MC solution and approximation errors). Other parameters: $\beta = 0, y_0 = 0.05, \nu = 0.4, c = 0.1, \rho = -0.1$.

The effect of c: the next parameter is c which controls the ATM levels of the smiles. As we fix $\nu = 0.4, \kappa = 0.05, \rho = -0.1$ and only vary c from 0.1 to 0.2, the curvature changes very little. In terms of the NIG's errors, we observe in figure 6.9 that their magnitude drops down as c increases but the difference is only clear when $T = 20$ or 30Y.

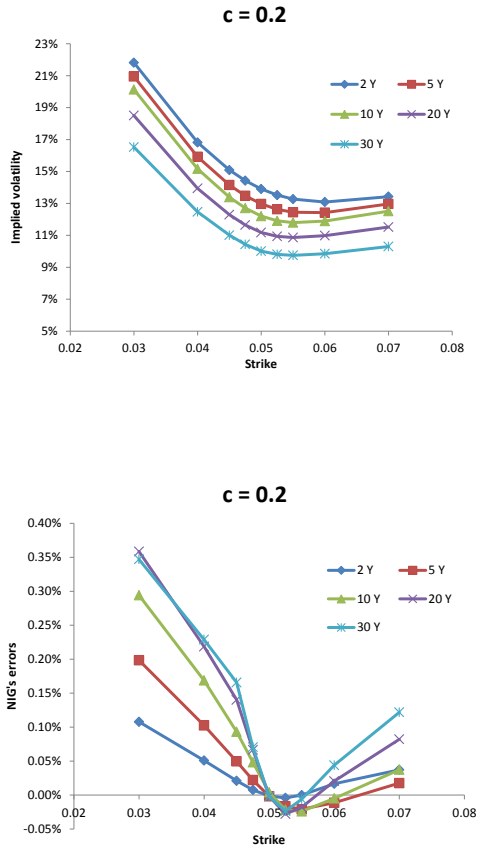


Figure 6.9: c-effect on the NIG approximation (MC solution and approximation errors). Other parameters: $\beta = 0, y_0 = 0.05, \nu = 0.4, \kappa = 0.05, \rho = -0.1$.

The effect of ρ: the last test is on the correlation parameter ρ. Note that it is often observed that ρ could be positive when $\beta = 0$ so we vary ρ from -0.1 to 0.1. As one can see from figure 6.10 as a comparison to figure 6.6, the smiles are of similar shapes but a bit more symmetric. The error curves for this case also seem to rotate anti-clockwise as a result of the smiles' rotation. This then leads to smaller error in the left wing but higher error in the right wing as ρ becomes more positive (or less negative).

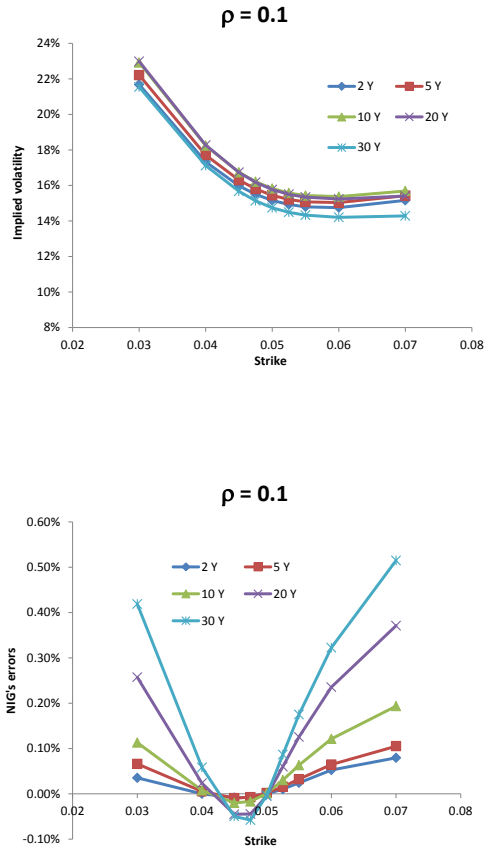


Figure 6.10: ρ-effect on the NIG approximation (MC solution and approximation errors). Other parameters: $\beta = 0, y_0 = 0.05, \nu = 0.4, \kappa = 0.05, c = 0.1$.

6.5.2 Log-Normal SABR-MR and DD-SABR-MR ($\beta \in (0, 1]$)

We continue our numerical investigation with other choices of β. Similar behaviour in the MC solutions as in the Normal SABR-MR model is observed when we vary parameters so we will not plot them here. In what follows, we display the NIG approximation's errors as we vary one parameter at a time when $\beta = 0.5, 1$. We consider the following typical parameter values and first investigate the effects of ν and κ which control the curvature of the smiles.

- $\beta = 0.5, y_0 = 0.05, a = 0.043, \rho = -0.2, c = 0.1$.
- $\beta = 1, y_0 = 0.05, a = 0.150, \rho = -0.5, c = 0.1$

The effect of ν: two values of $\nu = 0.25, 0.4$ are tested while $\kappa = 0.05$ and other parameters are held fixed as listed above. Figures 6.11 and 6.12 plot the effect of ν for the two models. At first glance, one can spot some immediate differences

between the shapes of error curves across models. In the Normal SABR-MR model, the NIG approximation tends to display a bit too much curvature (positive errors in both wings) whereas it shows more negative slope in the other models, i.e. implied volatilities are higher in the left wing and lower in the right wing. Note that the case $\beta = 1, \nu = 0.4$ yields quite large error for 30Y expiry in both left and right wings ($\approx 1\%$ at the -200 and 200 bp strike offset).

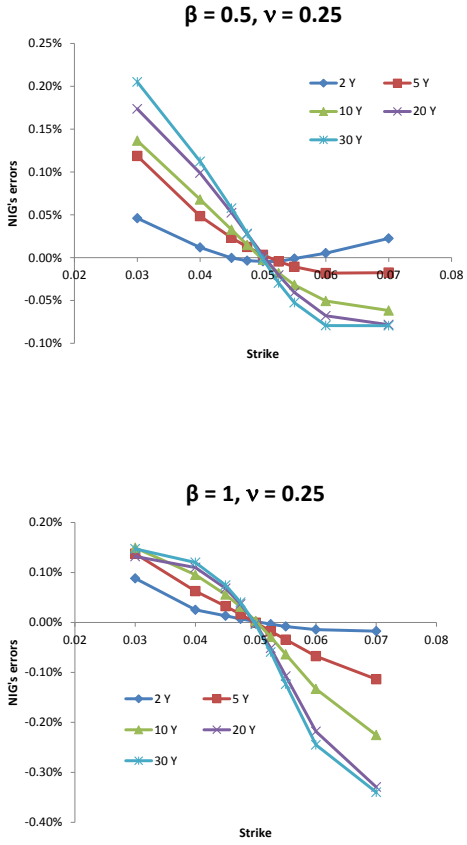


Figure 6.11: ν-effect on the NIG approximation.

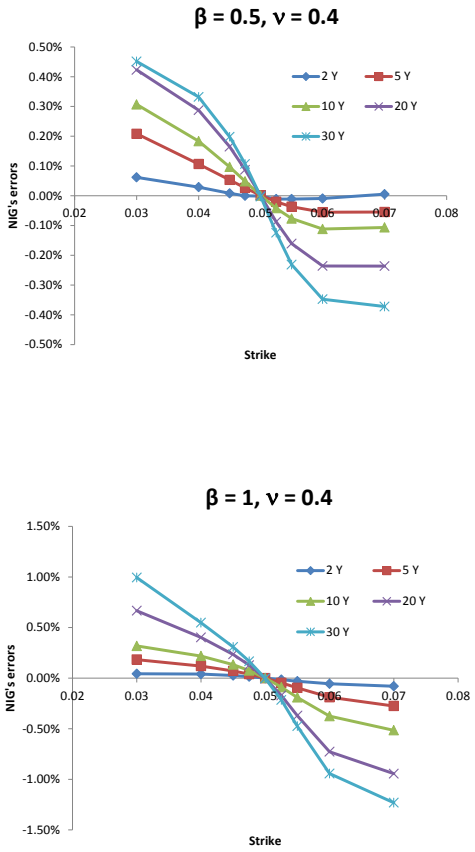


Figure 6.12: ν-effect on the NIG approximation.

The effect of κ: we fix $\nu = 0.4$ and increase the speed of mean reversion κ to 0.15. The κ-effect seems to be a lot clearer in these two models than the Normal SABR-MR model. Recall at this point that a larger κ makes the smiles flatten out faster as expiry increases. We observe that the left wing of the error curve gets lower and the right wing gets higher. For the $\beta = 1$ case, this effect seems to be the strongest. This implies that κ has a stronger effect on the NIG approximation in the left wing than in the right wing.

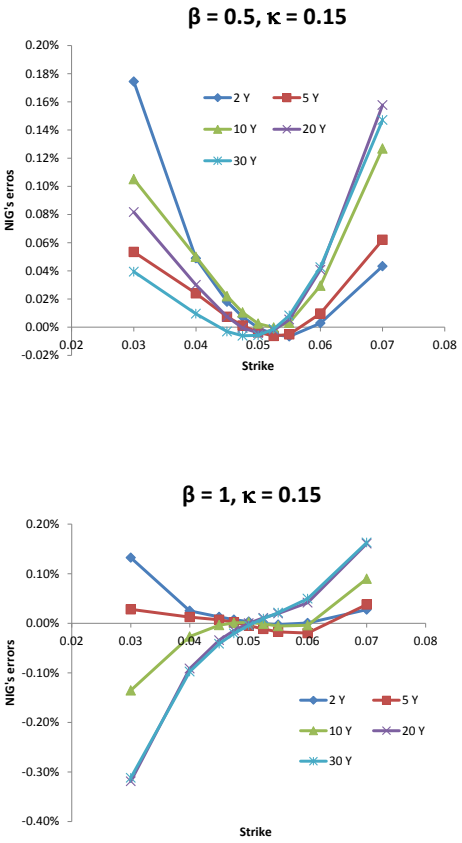


Figure 6.13: κ-effect on the NIG approximation.

The effects of c and ρ have also been investigated and similar results to those of the Normal SABR-MR model are obtained. We conclude up to this point that while c and ρ do not have big impacts on the quality of the NIG approximation, ν and κ are the main factors that determine its behavior. A combination of large $\nu = 0.4$ and low $\kappa = 0.05$ is the worst case in our numerical investigation so far for all β. Within this bound, the approximation performs reasonably well for all expiries. The worst scenario that occurs is when $T = 30$Y and $\beta = 1$. The errors for this case are the largest in magnitude for both left and right wings of the implied volatility curve.

6.5.3 Stress test

In this section, we further test the NIG approximation beyond the range of parameters used previously. The stress cases we consider here are extremely rare and very unlikely to happen in practice. These are also the volatility regimes when the NIG approximation starts to break down. We show for $\beta = 0$ only as other β's yield the

same conclusion.

- Stress test 1: $\nu = 0.6, \kappa = 0.05$ and other parameters $y_0 = 0.05, a = 0.0075, \rho = -0.1, c = 0.1$. This is an example of very large Volvol.
- Stress test 2: $\nu = 0.5, \kappa = 0.01$ and other parameters $y_0 = 0.05, a = 0.0075, \rho = -0.1, c = 0.1$. This is an example of large Volvol and low mean reversion speed.

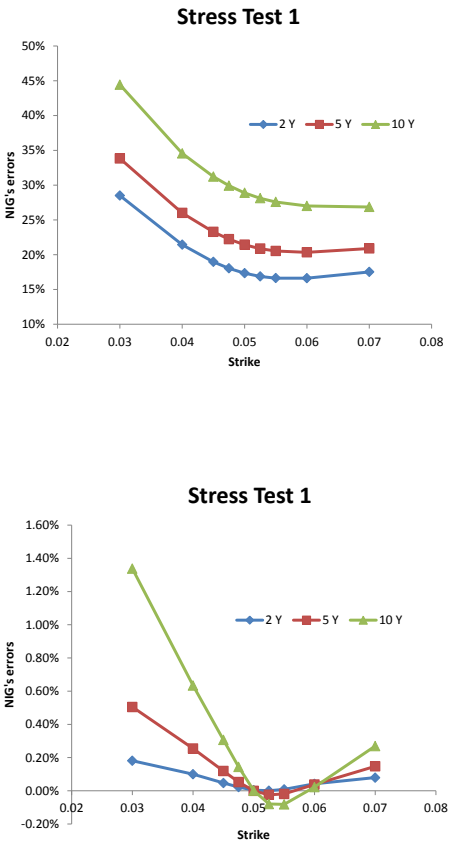


Figure 6.14: Stress test 1 for the NIG approximation (MC solution and approximation errors).

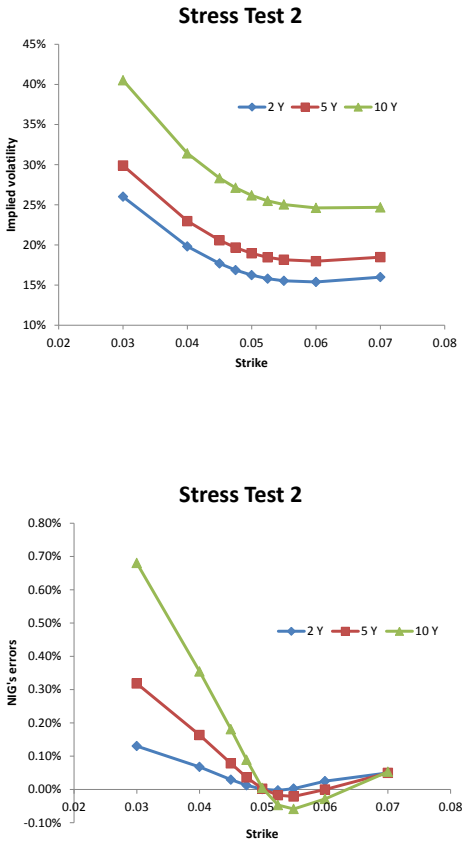


Figure 6.15: Stress test 2 for the NIG approximation (MC solution and approximation errors).

One can see from the above plots that the MC implied volatilities behave in an unusual way. The ATM level gets much higher as expiry increases and there also seems to be a lot more curvature for longer expiries. In both stress tests as displayed in figures 6.14 and 6.15, the NIG approximation only works up to 10Y expiry with large errors at the lowest strike ($\approx 1.4\%$ for the first test). For longer expiries, we need to roughly double the initial value of a to match up the ATM level. This is highly undesirable as this action tends to distort the intended distribution adversely. Specifically, even when we match the ATM levels the slope and curvature cannot be captured, i.e. too high in the left wing and too low in the right wing.

6.6 Conclusion

In this chapter, we have studied and set in context a possible extension of the SABR model which equips the volatility process with the mean reverting behaviour. In

Section 6.3, we examine the difference in dynamics of the two models. The results illustrate that mean reversion can have a stronger effect than Volvol on the dynamics of the volatility process through time. This is shown through a simple example of the forward start option. In this example, we find that the first two moments of the forward realized variance have the tendency to decrease with the mean reversion speed κ of the volatility process even though we calibrate the model properly to smiles of different expiries. Intuitively, we expect this effect causes the forward smile to flatten out more as κ increases and lead to that the SABR-MR model will underprice the forward start opions in the wings (strike ratios being away from 1) compared with the SABR model.

Using a similar approach to the probabilistic approximation in Chapter 5, we have derived a new approximation for the terminal distribution of the underlying in the SABR-MR model. We apply the approximation directly to price European options and obtain the corresponding implied volatilities. In our numerical study, we test the NIG approximation against the MC solution since we are not aware of any other current approximation for the SABR-MR model in the literature that works for a wide range of parameters. The results show that the NIG approximation can work well beyond the SABR model (as done in Chapter 5). The parameters we covered represent typical interest rate data through good and bad days of the market. The approximation only seems to break down when we worked with very rare parameters sets that represent extreme market scenarios. We want to emphasize again that while our focus tends to be on interest rate, the NIG approximation and the generic SABR-MR model are applicable to all other asset classes.

6.A Distribution of y_T under the Log-Normal SABR-MR model

The Log-Normal SABR-MR model is specified by the following SDEs

$$\begin{aligned} \mathrm{d}y_t &= I_t \exp(U_t) y_t \mathrm{d}W_t \qquad \beta \in [0,1], \\ \mathrm{d}U_t &= -\kappa U_t \mathrm{d}t + \nu \mathrm{d}Z_t \qquad \nu, \kappa > 0, \qquad U_0 = 0, \\ \mathrm{d}W_t \mathrm{d}Z_t &= \rho \mathrm{d}t \qquad \rho \in [-1,1], \end{aligned} \tag{6.17}$$

where $I_t = a \exp\{-c(T-t)\}$ is a deterministic function with $c > 0$. Write $W_t = \rho Z_t + \sqrt{1-\rho^2}\hat{W}_t$ where $Z \perp \hat{W}$. For the case $\beta = 1$ which we refer to as the

Log-Normal SABR-MR model, we have that

$$\begin{aligned}
\mathrm{d}y_t &= \rho I_t \exp(U_t) y_t \mathrm{d}Z_t + \sqrt{1-\rho^2} I_t \exp(U_t) y_t \mathrm{d}\hat{W}_t, \\
\mathrm{d}U_t &= -\kappa U_t \mathrm{d}t + \nu \mathrm{d}Z_t, \\
\mathrm{d}Z_t &= \frac{\mathrm{d}U_t + \kappa U_t \mathrm{d}t}{\nu}.
\end{aligned}$$

Let $F(t, U_t)$ be the anti-derivative of $I_t \exp(U_t)$ with respect to U_t, we have that $F(t, U_t) = I_t \exp(U_t)$. Applying Itô's lemma on $\ln y_t$, we have that

$$\begin{aligned}
& \mathrm{d} \ln y_t \\
= \; & \frac{1}{y_t} dy_t - \frac{1}{2(y_t)^2} \mathrm{d}[y]_t \\
= \; & I_t \exp(U_t) \mathrm{d}W_t - \frac{1}{2} (I_t)^2 \exp(2U_t) \mathrm{d}t \\
= \; & \rho I_t \exp(U_t) \mathrm{d}Z_t + \sqrt{1-\rho^2} I_t \exp(U_t) \mathrm{d}\hat{W}_t - \frac{1}{2} (I_t)^2 \exp(2U_t) \mathrm{d}t \\
= \; & \rho I_t \exp(U_t) \frac{\mathrm{d}U_t + \kappa U_t \mathrm{d}t}{\nu} + \sqrt{1-\rho^2} I_t \exp(U_t) d\hat{W}_t - \frac{1}{2} (I_t)^2 \exp(2U_t) \mathrm{d}t \\
= \; & \frac{\rho}{\nu} \mathrm{d}F(t, U_t) - \left(\frac{\rho\nu}{2} + \frac{c\rho}{\nu} \right) I_t \exp(U_t) \mathrm{d}t + \frac{\rho\kappa}{\nu} I_t \exp(U_t) U_t \mathrm{d}t \\
& + \sqrt{1-\rho^2} I_t \exp(U_t) \mathrm{d}\hat{W}_t - \frac{1}{2} (I_t)^2 \exp(2U_t) \mathrm{d}t \\
\Rightarrow \quad & \ln y_T \\
= \; & \ln y_0 + \frac{\rho}{\nu} (F(T, U_T) - F(0, U_0)) - \left(\frac{\rho\nu}{2} + \frac{c\rho}{\nu} \right) \int_0^T I_t \exp(U_t) \mathrm{d}t \\
& + \frac{\rho\kappa}{\nu} \int_0^T I_t \exp(U_t) U_t \mathrm{d}t - \frac{1}{2} \int_0^T (I_t)^2 \exp(2U_t) \mathrm{d}t + \sqrt{1-\rho^2} \int_0^T I_t \exp(U_t) \mathrm{d}\hat{W}_t,
\end{aligned}$$

where

$$\begin{aligned}
V_T &:= \int_0^T {I_t}^2 \exp(2U_t) \mathrm{d}t, \\
H_T &:= -\left(\frac{\nu}{2} + \frac{c}{\nu} \right) \int_0^T I_t \exp(U_t) \mathrm{d}t + \frac{\kappa}{\nu} \int_0^T I_t \exp(U_t) U_t \mathrm{d}t.
\end{aligned}$$

Similar to the proof in Appendix 5.A, by using the moment generating function (m.g.f) we have that

$$\ln y_T \triangleq \ln y_0 + \frac{\rho}{\nu} (F(T, U_T) - F(0, U_0)) + \rho H_T - \frac{1}{2} V_T + \sqrt{1-\rho^2} V_T^{\frac{1}{2}} G.$$

The same steps follow in both the Normal SABR-MR and DD-SABR-MR models to obtain (6.8) and (6.10) respectively.

6.B Distributions of y_{T_1} and y_{T_2} under the modified SABR-MR model in Section 6.3

Recall the modified SABR-MR model in Section 6.3. In this appendix, we want to show that when both ρ and $\beta = 0$ we have the following equations in distribution for $0 < T_1 < T_2$

$$\begin{aligned} y_{T_1} &\triangleq y_0 + \sqrt{V_{T_1}} G_1 \quad , \qquad V_{T_1} = \int_0^{T_1} \sigma_t^2 \mathrm{d}t, \\ y_{T_2} &\triangleq y_{T_1} + \sqrt{V_{T_2}^1} G_{12} \quad , \qquad V_{T_2}^1 = \int_{T_1}^{T_2} \sigma_t^2 \mathrm{d}t, \\ & G_1 \perp G_{12} \quad , \quad G_1, G_{12} \sim \mathcal{N}(0,1), \end{aligned}$$

and G_1 and G_{12} are both independent of $\{\sigma_t\}_{t\in[0,T_2]}$. Note that it is straightforward to apply the same proof from the previous appendix to obtain

$$\begin{aligned} y_{T_1} &\triangleq y_0 + \sqrt{V_{T_1}} G_1 \quad , \qquad V_{T_1} = \int_0^{T_1} \sigma_t^2 \mathrm{d}t, \\ y_{T_2} &\triangleq y_0 + \sqrt{V_{T_2}} G_2 \quad , \qquad V_{T_2} = \int_0^{T_2} \sigma_t^2 \mathrm{d}t, \\ & G_1 \perp G_2 \quad , \quad G_1, G_2 \sim \mathcal{N}(0,1). \end{aligned}$$

Hence, it suffices to prove

$$y_0 + \sqrt{V_{T_1}} G_1 + \sqrt{V_{T_2}^1} G_{12} \triangleq y_0 + \sqrt{V_{T_2}} G_2,$$

where G_1,G_{12} and G_2 are pairwise independent.

Recall from the previous appendix that $\mathcal{F}_T = \sigma(Z_u : 0 \le u \le T)$. It is clear that

$$\begin{aligned} \sqrt{V_{T_1}} G_1 | \mathcal{F}_{T_2} &\sim \mathcal{N}(0, V_{T_1}), \\ \sqrt{V_{T_2}} G_2 | \mathcal{F}_{T_2} &\sim \mathcal{N}(0, V_{T_2}), \\ \sqrt{V_{T_2}^1} G_{12} | \mathcal{F}_{T_2} &\sim \mathcal{N}(0, V_{T_2}^1). \end{aligned}$$

Let $X_{T_2} := \sqrt{V_{T_1}} G_1 + \sqrt{V_{T_2}^1} G_{12}$. For the first conditional moment of $X_{T_2}|\mathcal{F}_{T_2}$, we have that

$$\begin{aligned} \mathbb{E}(X_{T_2}|\mathcal{F}_{T_2}) &= \mathbb{E}(\sqrt{V_{T_1}} G_1 | \mathcal{F}_{T_2}) + \mathbb{E}(\sqrt{V_{T_2}^1} G_{12} | \mathcal{F}_{T_2}) \\ &= \sqrt{V_{T_1}} \mathbb{E}(G_1 | \mathcal{F}_{T_2}) + \sqrt{V_{T_2}^1} \mathbb{E}(G_{12} | \mathcal{F}_{T_2}) \\ &= 0, \end{aligned}$$

since G_1 and G_{12} are indepdendent of $\mathcal{F}_{T_2}$. Similarly, for the second conditional moment of $X_{T_2}|\mathcal{F}_{T_2}$ we have that

$$\begin{aligned}
\mathbb{E}(X_{T_2}^2|\mathcal{F}_{T_2}) &= \mathbb{E}(V_{T_1}G_1^2|\mathcal{F}_{T_2}) + \mathbb{E}(V_{T_2}^1 G_{12}^2|\mathcal{F}_{T_2}) + 2\mathbb{E}(G_1 G_{12}\sqrt{V_{T_1}V_{T_2}^1}|\mathcal{F}_{T_2}) \\
&= V_{T_1}\mathbb{E}(G_1^2|\mathcal{F}_{T_2}) + V_{T_2}^1\mathbb{E}(G_{12}^2|\mathcal{F}_{T_2}) + 2\sqrt{V_{T_1}V_{T_2}^1}\mathbb{E}(G_1G_{12}|\mathcal{F}_{T_2}) \\
&= V_{T_1} + V_{T_2}^1 + 2\sqrt{V_{T_1}V_{T_2}^1}\mathbb{E}(G_1G_{12}) \\
&= V_{T_2}.
\end{aligned}$$

The last line follows since $G_1 \perp G_{12}$ and $V_{T_1} + V_{T_2}^1 = V_{T_2}$. Similar to the previous appendix, by considering the m.g.f of X_{T_2} we have that

$$\begin{aligned}
\mathbb{E}(e^{sX_{T_2}}) &= \mathbb{E}(\mathbb{E}(e^{sX_{T_2}}|\mathcal{F}_{T_2})) \\
&= \mathbb{E}\left(e^{\frac{1}{2}s^2V_{T_2}}\right) \\
&= \mathbb{E}(\mathbb{E}(e^{s\sqrt{V_{T_2}}G_2}|\mathcal{F}_{T_2})) \\
&= \mathbb{E}(e^{s\sqrt{V_{T_2}}G_2}).
\end{aligned}$$

Hence, we have that $X_{T_2} \triangleq \sqrt{V_{T_2}}G_2$ or equivalently $y_{T_2} \triangleq y_{T_1} + \sqrt{V_{T_2}^1}G_{12}$.

6.C Proof of Proposition 3

In order to calculate $\mathbb{E}(V_T|U_T = u^*)$, $\mathbb{E}(V_T^2|U_T = u^*)$ and $\mathbb{E}(H_T|U_T = u^*)$, it is sufficient to look at the distribution of y_T conditioned on the end point U_T which is fully specified by the so called OU bridge. Note that U_t is a Gaussian process with law

$$\begin{aligned}
U_t &\sim \mathcal{N}(m(t), R(t,t)), \\
m(t) &= 0, \\
R(t,t) &= \frac{\nu^2}{2\kappa}[1 - \exp(-2\kappa t)].
\end{aligned}$$

Furthermore, it has the covariance function for $0 \leq t, s \leq T$:

$$R(t,s) = \mathrm{Cov}(U_t, U_s) = \frac{\nu^2}{2\kappa}[\exp\{-\kappa|t-s|\} - \exp\{-\kappa(t+s)\}].$$

For ease of exposition, by $R(t)$ we mean the variance $R(t,t)$. An OU bridge is an OU process that is "tied down" at time T to have some specific value. We denote the OU bridge of the corresponding OU process U by U^{T,u^*} from $(0,0)$ to (T,u^*) which

admits the following non-anticipative representation (see Gasbarra et al. [2007])

$$U_t^{T,u^*} \triangleq U_t - \frac{R(t,T)}{R(T)} U_T + u^* \frac{R(t,T)}{R(T)}. \tag{6.18}$$

Note that this representation coincides with the Brownian bridge representation when U is a Brownian motion. It then follows for $0 \leq t, s \leq T$ that, U^{T,u^*} is Gaussian with:

$$\begin{aligned} m^{T,u^*}(t) &= \mathbb{E}(U_t^{T,u^*}) = u^* \frac{R(t,T)}{R(T)}, & (6.19) \\ R^{T,u^*}(t,s) &= \mathrm{Cov}(U_t^{T,u^*}, U_s^{T,u^*}) = R(t,s) - \frac{R(t,T)R(s,T)}{R(T)}, & (6.20) \end{aligned}$$

and $R^{T,u^*}(t) = R^{T,u^*}(t,t)$. As this OU bridge is also a Gaussian process with the above mean and covariance function, we have that:

$$\exp(U_t^{T,u^*}) \sim \mathcal{LN}(m^{T,u^*}(t), R^{T,u^*}(t)). \tag{6.21}$$

6.C.1 First conditional moments of H_T and V_T

We now evaluate the first conditional moments of H_T and V_T whose forms are given in (6.12) and (6.11) in Section 6.4 in the main chapter respectively. Conditioning on $U_T = u^*$, we have that

$$\begin{aligned} & \mathbb{E}(H_T | U_T = u^*) \\ = & -\left(\frac{\nu}{2} + \frac{c}{\nu}\right) \int_0^T I_t \mathbb{E}\{\exp(U_t^{T,u^*})\} \mathrm{d}t + \frac{\kappa}{\nu} \int_0^T I_t \mathbb{E}\{\exp(U_t^{T,u^*}) U_t^{T,u^*}\} \mathrm{d}t \\ = & -\left(\frac{\nu}{2} + \frac{c}{\nu}\right) \underbrace{\int_0^T I_t \exp\left(m^{T,u^*}(t) + \frac{R^{T,u^*}(t)}{2}\right) \mathrm{d}t}_{H_1(u^*)} \qquad (6.22) \\ & + \frac{\kappa}{\nu} \underbrace{\int_0^T I_t (m^{T,u^*}(t) + R^{T,u^*}(t)) \exp\left(m^{T,u^*}(t) + \frac{R^{T,u^*}(t)}{2}\right) \mathrm{d}t}_{H_2(u^*)}, \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}(V_T | U_T = u^*) &= \int_0^T (I_t)^2 \mathbb{E}\{\exp(2U_t^{T,u^*})\} \mathrm{d}t \\ &= \int_0^T (I_t)^2 \exp\left(2m^{T,u^*}(t) + 2R^{T,u^*}(t)\right) \mathrm{d}t. \end{aligned} \tag{6.23}$$

Expand further the above integrals we have that

$$\begin{aligned}
H_1(u^*) &= \int_0^T I_t \exp\left(u^* \frac{R(t,T)}{R(T)} + \frac{R(t) - \frac{R^2(t,T)}{R(T)}}{2}\right) \mathrm{d}t, \\
H_2(u^*) &= \int_0^T I_t \left(u^* \frac{R(t,T)}{R(T)} + R(t) - \frac{R^2(t,T)}{R(T)}\right) \\
&\quad \times \exp\left(u^* \frac{R(t,T)}{R(T)} + \frac{R(t) - \frac{R^2(t,T)}{R(T)}}{2}\right) \mathrm{d}t, \\
\mathbb{E}(V_T | U_T = u^*) &= \int_0^T (I_t)^2 \exp\left(2u^* \frac{R(t,T)}{R(T)} + 2R(t) - 2\frac{R^2(t,T)}{R(T)}\right) \mathrm{d}t.
\end{aligned}$$

6.C.2 Second conditional moment of V_T

We now evaluate the second conditional moment of V_T.

$$\begin{aligned}
&\mathbb{E}(V_T^2 \mid U_T = u^*) \\
= \;& 2\mathbb{E}\left(\int_0^T (I_t)^2 e^{2U_t} \int_0^t (I_s)^2 e^{2U_s} \mathrm{d}s\mathrm{d}t | U_T = u^*\right) \\
= \;& 2\mathbb{E}\left(\int_0^T (I_t)^2 \int_0^t (I_s)^2 e^{2U_t + 2U_s} \mathrm{d}s\mathrm{d}t | U_T = u^*\right) \\
= \;& 2\int_0^T (I_t)^2 \int_0^t (I_s)^2 \mathbb{E}\left(e^{2U_t^{T,u^*} + 2U_s^{T,u^*}}\right) \mathrm{d}s\mathrm{d}t \\
= \;& 2\int_0^T (I_t)^2 \int_0^t (I_s)^2 \exp(2m^{T,u^*}(t) + 2m^{T,u^*}(s) + 2R^{T,u^*}(t) + 2R^{T,u^*}(s) + 4R^{T,u^*}(t,s)) \mathrm{d}s\mathrm{d}t \\
= \;& 2\int_0^T (I_t)^2 \int_0^t (I_s)^2 \exp\left(2\left(u^* \frac{R(t,T)}{R(T)} + u^* \frac{R(s,T)}{R(T)}\right)\right) \\
& \times \exp\left(2R(t) - 2\frac{R^2(t,T)}{R(T)} + 2R(s) - 2\frac{R^2(s,T)}{R(T)} + 4R(t,s) - 4\frac{R(t,T)R(s,T)}{R(T)}\right) \mathrm{d}s\mathrm{d}t.
\end{aligned}$$

Chapter 7

Hedging European options with stochastic volatility models

7.1 Objectives

In the previous chapters, we have focused on the problem of pricing European options within the SABR and SABR-MR stochastic volatility models in quite some detail. As hinted in Chapter 6, although the SABR-MR model is more flexible than the SABR model they are qualitatively very similar in terms of pricing European options. In order to gain better understanding and intuition of the two models, this chapter investigates further their hedging properties. Note that the comparison of these models from a hedging perspective has not been covered previously in the literature.

Recall the notation y_t and σ_t for the underlying and the volatility process respectively for $t \in [0, T]$. Roughly speaking, a model that is capable of calibrating to the market smile at time T will get right the following marginal distribution

$$(y_T | y_0 = y, \sigma_0 = \sigma). \tag{7.1}$$

While this seems to be the case for both the SABR and SABR-MR models, it has not yet been investigated whether they will also agree on

$$(y_T | y_0, \sigma_0). \tag{7.2}$$

In other words, suppose the two models have been calibrated to (7.1) for $y_0 = y$ and $\sigma_0 = \sigma$, the question is: will they continue to agree if we perturb the state variables by a small amount, e.g. $y_0 = y + \epsilon$ and/or $\sigma_0 = \sigma + \epsilon$ for $\epsilon > 0$? An example of such disagreement was done in Hagan et al. [2002] for the SABR model and a simple local volatility model. The authors perturb y_0 with σ_0 held fixed and

look at the shift of implied volatility smiles of the two models. The fact that they move in opposite directions implies that (7.2) is not the same for the two models. The investigation in Hagan et al. [2002] essentially examines the sensitivities of the models with respect to the state variables.

We want to mention at this point that studying the previous question will not give us a full picture about the hedging problem. In fact, we will only be able to learn about the static behaviour of the volatility smile, i.e. observe different versions of today's smile but not let time vary. It is actually more important to appreciate the smile dynamics that is generated by the model for a full understanding of hedging. The hedge ratios that are based on considering the model dynamics have been mentioned previously in the literature, e.g. Bartlett [2006], Rebonato et al. [2009] and Andersen and Piterbarg [2010]. In Section 7.2, we want to separate these different concepts and explain what really matters in practice when we hedge with the presence of stochastic volatility. We then apply these concepts to the SABR and SABR-MR models to assess whether the two models will differ from this perspective.

7.2 Hedging with stochastic volatility: theory and practice

Before assessing the hedging properties of the SABR and SABR-MR models, we first review different hedging concepts in detail.

7.2.1 The theoretical concept: from deterministic to stochastic volatility

We first denote by $C_0(K, y_0, \sigma_0, T)$ or just C_0 (for abbreviation) today's price of a European call option with strike K, initial underlying y_0, option maturity T and initial volatility σ_0. The definitions of "pure" delta δ_1 and vega v_1 today can be defined as follows

$$\delta_1 = \frac{\partial C_0(K, y_0, \sigma_0, T)}{\partial y_0}, \tag{7.3}$$

$$v_1 = \frac{\partial C_0(K, y_0, \sigma_0, T)}{\partial \sigma_0}. \tag{7.4}$$

These model sensitivities are interpreted as the rates of change in price as y_0 or σ_0 changes assuming that the other is kept fixed. The computation of the pure delta and vega is typically done numerically by finite difference, e.g. δ_1 can be approximated as

$$\delta_1 \approx \frac{C_0(K, y_0 + \epsilon, \sigma_0, T) - C_0(K, y_0, \sigma_0, T)}{\epsilon},$$

where the perturbation size $\epsilon > 0$ is sufficiently small. In the Black-Scholes economy and hence a complete market, δ_1 can be interpreted as the amount of holdings of y_0 in the replicating portfolio for this European option. This means that if we acquire a hedging portfolio consisting of this European option and a short δ_1 position of y_0, the growth of this portfolio regardless of the underlying will be deterministic (or locally riskless) over $[0, \Delta t]$ where Δt denotes a short period of time. This definition of delta coincides with what we defined in Part I of the thesis. The reason for why this result holds true is that volatility is assumed to be constant or deterministic over $[0, \Delta t]$ in the Black-Scholes economy. This will no longer be true if we relax this assumption and let volatility to be stochastic.

In the stochastic volatility setting, it is clear that the market is no longer complete if one just purely adopts the delta hedging with acquiring only positions in y_0. This is because the stochastic volatility process σ is driven by a second Brownian motion Z (correlated with W which drives y). However, if the volatility σ is an available traded asset in the market they can be included in the hedging portfolio and the market will be complete. See Chapter 1, Hunt and Kennedy [2004] for a discussion on market completeness. See also Lemma 8.9.1 in Andersen and Piterbarg [2010] for the theoretical construction of a locally riskless portfolio in the stochastic volatility setting by taking positions in two traded assets that depend on both the underlying and volatility.

An important point about the consequence of introducing stochastic volatility that we want to make here is that the pure delta δ_1 cannot be used naively as a hedge ratio to construct a locally riskless portfolio as in the Black-Scholes economy. We do not go further into how to construct a locally riskless portfolio or to complete the market, but we will extend further the idea of delta hedging that employs only the underlying asset. It is useful to look at the *smile dynamics* generated by general stochastic volatility models for a better understanding of this subject. A rigorous treatment of the smile dynamics is addressed in Chapter 12 of Rebonato et al. [2009] or Volume 1 of Andersen and Piterbarg [2010] (page 348 to 352). In Andersen and Piterbarg [2010], for numerical experiments the authors move calendar time forward to some arbitrary value t and examine how the smile might look for several future levels of y_t. A similar investigation was done in Mercurio and Morini [2008] but for a very short period of time Δt rather than an arbitrary t. As our purpose is to consider a form of delta that has a meaningful financial interpretation, we first want to assess how a small movement of the underlying, say $\Delta y > 0$, might contribute to the overall change in price of a European option after a short period of time Δt. We then construct an appropriate portfolio for the European option that is not too exposed to this movement of the underlying. Hence, a sensible way of viewing the

delta following this motivation is

$$\begin{aligned} \delta_2 &\approx \frac{\Delta C}{\Delta y}, \\ \Delta C &= C_0(K, y_0 + \Delta y, \sigma_{0+\Delta t}, T) - C_0(K, y_0, \sigma_0, T) \\ \Delta y &= y_{0+\Delta t} - y_0, \end{aligned} \tag{7.5}$$

where Δt and the perturbation size Δy are both sufficiently small and positive. Here, δ_2 reflects the rate of change in the European option price as the underlying evolves. The evaluation of δ_2 is not intuitively clear at first glance. To be more specific, the quantity $C_0(K, y_0+\Delta y, \sigma_{0+\Delta t}, T)$ will need to be valued at the volatility level $\sigma_{0+\Delta t}$ which is the future volatility level. Note that δ_1 and δ_2 will be identical if $\sigma_{0+\Delta t} = \sigma_0$. The problem is that the volatility process is not observable in the market so it is impossible to know the exact size of perturbation for volatility, i.e. $\Delta\sigma = \sigma_{0+\Delta t} - \sigma_0$, given that we know Δy. Hence, δ_2 cannot be observed without knowledge of data. From the model perspective, the only power we have is to estimate $\Delta\sigma$ given the known Δy via the known SDE formulation. A sensible thing to do is to consider the conditional expectation $\mathbb{E}(\Delta\sigma|\Delta y)$. This idea was considered in a few references in the literature, e.g. Mercurio and Morini [2008] and Andersen and Piterbarg [2010]. This conditional expectation has a clear meaning in terms of real data. If the model best describes market data, $\mathbb{E}(\Delta\sigma|\Delta y)$ will be the average of all possible real world realizations of the future change in volatility given an anticipated future change in the underlying process. We can evaluate this conditional expectation by decomposing the SDE for volatility into two orthogonal Brownian motions W and $W^\perp$. In case of the SABR model, if we write the Brownian motion that drives σ as $Z_t = \rho W_t + \sqrt{1-\rho^2} W_t^\perp$ we will have that

$$\begin{aligned} \mathrm{d}\sigma_t &= \nu\sigma_t \mathrm{d}Z_t \\ &= \nu\sigma_t(\rho \mathrm{d}W_t + \sqrt{1-\rho^2}\mathrm{d}W_t^\perp) \\ &= \nu\rho\frac{\mathrm{d}y_t}{y_t^\beta} + \sqrt{1-\rho^2}\nu\sigma_t \mathrm{d}W_t^\perp. \end{aligned}$$

As we know Δy, following the above SDE we have that

$$\mathbb{E}(\Delta\sigma|\Delta y) = \nu\rho\frac{\Delta y}{y_0^\beta} + \sqrt{1-\rho^2}\nu\sigma_0\mathbb{E}(\Delta W^\perp|\Delta y).$$

Since the process y and W are independent of $W^\perp$, the second conditional expectation in the above equation is just zero. We then arrive at

$$\mathbb{E}(\Delta\sigma|\Delta y) = \nu\rho\frac{\Delta y}{y_0^\beta}.$$

Since we are working with a delta that is an estimate (or average) of δ_2, we will denote it by $\widetilde{\delta}_2$. Putting pieces together, we have the following

$$\widetilde{\delta}_2 \approx \frac{C_0(K, y_0 + \Delta y, \sigma_0 + \mathbb{E}(\Delta\sigma|\Delta y), T) - C_0(K, y_0, \sigma_0, T)}{\Delta y}.$$

Applying the first order Taylor expansion on the price function, we have that

$$\begin{aligned} & C_0(K, y_0 + \Delta y, \sigma_0 + \mathbb{E}(\Delta\sigma|\Delta y), T) \\ \approx\ & C_0(K, y_0, \sigma_0, T) + \Delta y \frac{\partial C_0(K, y_0, \sigma_0, T)}{\partial y_0} + \mathbb{E}(\Delta\sigma|\Delta y) \frac{\partial C_0(K, y_0, \sigma_0, T)}{\partial \sigma_0} \\ \approx\ & C_0(K, y_0, \sigma_0, T) + \delta_1 \Delta y + \nu\rho \frac{\Delta y}{y_0^\beta} v_1. \end{aligned}$$

This then implies that

$$\widetilde{\delta}_2 \approx \delta_1 + \frac{\nu\rho}{y_0^\beta} v_1.$$

The right hand side of the above equation first appeared in Bartlett [2006] so we will refer to $\widetilde{\delta}_2$ as the Bartlett delta. Following the same approach as above, we obtain exactly the same formula for $\widetilde{\delta}_2$ of the DD-SABR model.

In case of the (DD-)SABR-MR model, the argument follows similarly. We first recall from Chapter 6 that $\sigma_t = a\exp(-c(T-t))\exp(U_t)$ and $\mathrm{d}U_t = -\kappa U_t \mathrm{d}t + \nu \mathrm{d}Z_t$ where a, c, ν and κ are all positive constants. By Itô's formula, we have that

$$\begin{aligned} \mathrm{d}\sigma_t &= \sigma_t \mathrm{d}U_t + \frac{1}{2}\sigma_t \mathrm{d}[U]_t + c\sigma_t \mathrm{d}t \\ &= \sigma_t(-\kappa U_t \mathrm{d}t + \nu \mathrm{d}Z_t) + \frac{1}{2}\sigma_t \nu^2 \mathrm{d}t + c\sigma_t \mathrm{d}t \\ &= \nu\sigma_t \mathrm{d}Z_t + \sigma_t(-\kappa U_t + \frac{1}{2}\nu^2 + c)\mathrm{d}t. \end{aligned}$$

Again, by decomposing the Brownian motion Z into two orthogonal counterparts and ignoring the finite variation part we have that

$$\mathbb{E}(\Delta\sigma|\Delta y) \approx \nu\rho \frac{\Delta y}{y_0^\beta}.$$

We then arrive at the same form of $\widetilde{\delta}_2$ for the (DD-)SABR-MR model.

The above discussion gives the Bartlett delta a clear financial interpretation which does not follow from the original presentation in Bartlett [2006]. It reflects the change in price given a change in the underlying and a predicted average (conditional) change in the volatility. In terms of hedging, the hedge construction that adopts the Bartlett delta turns out to have some practical relevance. Suppose we

wish to delta hedge a European option with today's value C_0 by going short δ_{mv}[1] of the underlyings, i.e. the hedging portfolio has today's value $\Pi_0 = C_0 - \delta_{\text{mv}} y_0$. Since a perfect replication of this European option is not available in the stochastic volatility setting, the best that we can hope for is that the value of the hedging portfolio (given that the model is reasonably realistic) will not vary too much over a short period of time, say $[0, \Delta t]$. Practically, this means that the variance of the Profit & Loss of the hedging portfolio will be small. Equivalently in mathematical terms, we would want to have that $\Delta\Pi = \Pi_{0+\Delta t} - \Pi_0$ has the minimum variance. As shown explicitly in Andersen and Piterbarg [2010] (Lemma 8.9.4, Volume 1) and Rebonato et al. [2009] (page 221-229) for general stochastic volatility models and the SABR model in particular, the minimum variance portfolio can be achieved by letting δ_{mv} to be the same as the Bartlett delta, i.e.

$$\delta_{\text{mv}} = \delta_1 + \frac{\nu\rho}{y_0^{\beta}} v_1.$$

The analysis of vega could be proceeded along the same lines as the above, e.g. see Bartlett [2006]. While the underlying y is directly observable in the market, the volatility process σ is not. As a result, one can not easily observe the change in price as σ changes. In practice, one tends to assess the the change in price with respect to some other market quantities, e.g. Black ATM implied volatility or Normal ATM implied volatility. We will address this issue in Section 7.2.2 where we consider more practical forms of the Greeks.

7.2.1.1 Numerical Examples

We recall that our primary purpose of this chapter is to assess if the (DD-)SABR and (DD-)SABR-MR models have different hedging properties. In what follows, we will look at both the pure delta and the Bartlett delta numerically. We have compared both models in terms of these deltas for a particular example and the results are comparable. This concludes that mean reversion does not influence the deltas a lot and the two models continue to be similar.

Since the numerical example that we are going to consider shows that the two models yield very similar deltas, we will only display the results for the (DD-)SABR model. For a range of calibrated parameters on 27/10/2007 (table 7.1), we compute the pure delta (δ_1) and the Bartlett delta ($\widetilde{\delta}_2$) for swaptions with different strikes and different expiries. The results are displayed in figures 7.1 and 7.2 below.

[1] The subscript is the abbreviation for minimum variance.

Expiry	β	σ_0	ν	ρ	β	σ_0	ν	ρ
2Y	0	0.0057	0.309	0.060	0.5	0.026	0.326	-0.160
5Y	0	0.0055	0.275	0.003	0.5	0.025	0.297	-0.220
10Y	0	0.0052	0.243	0.027	0.5	0.023	0.263	-0.205
20Y	0	0.0044	0.203	0.035	0.5	0.020	0.223	-0.219
30Y	0	0.0038	0.183	0.041	0.5	0.018	0.203	-0.233

Table 7.1: Calibrated parameters for the Normal SABR and DD-SABR models for different expiries.

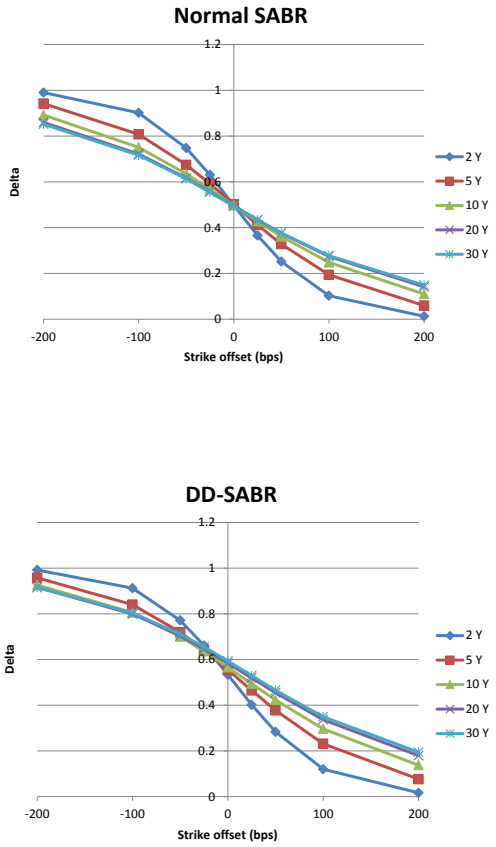


Figure 7.1: Pure delta δ_1 under the Normal SABR and DD-SABR models.

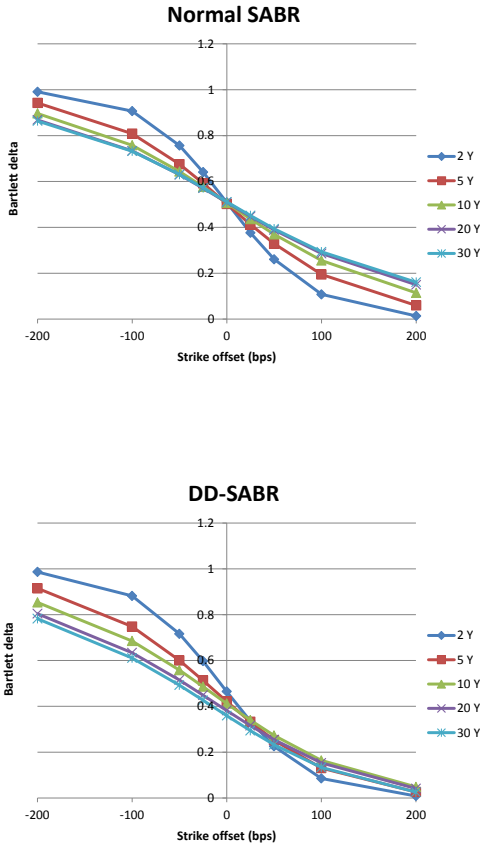


Figure 7.2: Bartlett delta $\widetilde{\delta}_2$ under the Normal SABR and DD-SABR models.

For short expiries ($< 10Y$), the Bartlett delta $\widetilde{\delta}_2$ seems to be independent of the choice of β and ρ while the pure delta δ_1 can increase significantly as β increases or equivalently ρ becomes more negative. Note that the clearest effect is seen around the ATM region. The Bartlett delta for $\beta = 0$ case is still very similar to the pure delta while the extra term $\frac{\nu\rho}{y_0^{\beta}} v_1$ in the $\beta = 0.5$ case seems to pull the pure delta down significantly around the ATM region. This observation is confirmed by earlier results in Bartlett [2006] and Andersen and Piterbarg [2010] for a 5Yx5Y swaption example. However in contrast to their results, we find that a stronger effect is observed for longer expiries which has not been noted in the literature. That is the longer the expiry is the more the Bartlett delta is affected for $\beta = 0.5$. This is because the extra term $\frac{\nu\rho}{y_0^{\beta}} v_1$ could become a lot more negative as expiry increases to 20Y or 30Y. Consequently, this leads to the result that the Bartlett delta can decrease substantially as β increases for long expiries.

7.2.2 Practical delta and vega

As hinted in the previous section, the pure delta and the Bartlett delta have very specific theoretical interpretations. While the Bartlett delta appears to have some positive impact on the delta hedge of a European option due to the minimum variance portfolio it imposes, this hedging strategy remains to be tested empirically. The reason for this is the following. The minimum variance hedge argument is based on a completely theoretical framework which relies on the model's initial parametrization (the set of parameters required to calibrate to market smiles). This means that if the market in reality behaves differently from the dynamics predicted by the model for various reasons, we will probably end up with a wrong delta figure and the minimum variance hedge will not be financially meaningful. This is why practitioners tend to base the delta hedge more on historical data rather than completely rely on the chosen model. In what follows, we describe a more practical way of calculating the delta and vega.

In a typical example of a vanilla European option trading desk, practitioners use stochastic volatility models on a daily basis to calibrate to market data. During each trading day, some parameters are found to be stable and some are not. This encourages us to hedge the risks involving model parameters as well as model state variables. Depending on the nature of the market, practitioners can choose different ways to parametrize their models and it is this difference that causes the presence of different hedge ratios. The type of hedging that involves neutralizing the risks with respect to model parameters is usually referred to as out-of-model hedging. See Rebonato [2004] and Rebonato et al. [2009] for discussions on various types of hedging including both in-model and out-of-model hedging.

Recall from Chapter 4 that the Normal implied volatility is the volatility parameter that we plug into the Bachelier formula to recover the market price. From talking to practitioners, it is not unusual in some market that the Normal ATM implied volatility $\widetilde{\sigma}_{\mathrm{N,ATM}}(y_0)$ is quite stable day to day. This means that although this quantity could vary daily, this should not be treated as a systematic change from the movement of the underlying asset. Hence, $\widetilde{\sigma}_{\mathrm{N,ATM}}(y_0)$ and y_0 should be treated as independent market variables/input to the model. We illustrate this intuition in the following example. Suppose we choose a fixed $\beta > 0$ for the SABR model to calibrate to data. This requires us to also fix an appropriate value of ρ to capture the right skew level. Roughly speaking, if the Volvol ν is low indicating a small effect of stochastic volatility on the smile, we will have that the average level of the Normal ATM implied volatility $\widetilde{\sigma}_{\mathrm{N,ATM}}(y_0)$ will be approximately $\sigma_0 y_0^{\beta}$ (see Chapter 4 for a rough intuition of the Normal implied volatility). This means that if y_0 moves, we will observe a systematic change in $\widetilde{\sigma}_{\mathrm{N,ATM}}(y_0)$ which is not desired in our particular context. To prevent this from happening, we have to adjust the

model volatility level σ_0 so that $\widetilde{\sigma}_{\mathrm{N,ATM}}(y_0)$ will remain unchanged. Effectively, the initial volatility σ_0 should be of the form $\sigma_0(\widetilde{\sigma}_{\mathrm{N,ATM}}(y_0), y_0)$.

Bearing this intuition in mind, we now consider a new form of delta

$$\begin{aligned}\delta_{\text{practice}} &= \frac{\partial C_0(K, y_0, \sigma_0, T)}{\partial y_0} + \frac{\partial C_0(K, y_0, \sigma_0, T)}{\partial \sigma_0}\frac{\partial \sigma_0(\widetilde{\sigma}_{\mathrm{N,ATM}}(y_0), y_0)}{\partial y_0} \\ &= \delta_1 + \frac{\partial \sigma_0(\widetilde{\sigma}_{\mathrm{N,ATM}}(y_0), y_0)}{\partial y_0} v_1. \end{aligned} \tag{7.6}$$

This delta risk is now the risk with respect to changes in y_0 with $\widetilde{\sigma}_{\mathrm{N,ATM}}(y_0)$ held fixed. The partial derivative in the last term is just the change in σ_0 needed to keep $\widetilde{\sigma}_{\mathrm{N,ATM}}(y_0)$ constant while y_0 changes. We note that the motivation for this form of delta is very different from the Bartlett delta. Whilst the Bartlett delta arises from the effect of the correlation parameter ρ, this practical delta that we are considering tends to be an immediate adjustment to account for the β-effect.

The story of vega is slightly simpler than the delta. As discussed in the previous section, in practice it may be more relevant to consider the change in some market volatility quantity rather than the model volatility (initial volatility σ_0). If one uses the Normal ATM implied volatility $\widetilde{\sigma}_{\mathrm{N,ATM}}(y_0)$ as the model parameter, it seems to make more sense to calculate the vega as the model sensitivity to this market parameter. Since $\widetilde{\sigma}_{\mathrm{N,ATM}}(y_0)$ and y_0 are assumed to be two independent variables of the model, a change in $\widetilde{\sigma}_{\mathrm{N,ATM}}(y_0)$ will have no effect on y_0. Hence, we have the following practical vega which is basically the scaled pure vega

$$v_{\text{practice}} = \frac{\partial C_0(K, y_0, \sigma_0, T)}{\partial \widetilde{\sigma}_{\mathrm{N,ATM}}(y_0)} = \frac{\partial C_0(K, y_0, \sigma_0, T)}{\partial \sigma_0} \Big/ \frac{\partial \widetilde{\sigma}_{\mathrm{N,ATM}}(y_0)}{\partial \sigma_0}. \tag{7.7}$$

7.2.2.1 Numerical Examples

In this section, we will look at the practical delta δ_{practice} and the practical vega v_{practice} numerically. Having discussed in detail their meanings and financial interpretations, we now compute them within the SABR and SABR-MR models to assess if different model formulations have a significant impact on these Greeks. We again note that this investigation has not been done in the literature. In order to have a fair comparison, we will choose parameters such that European swaption prices produced by the two models coincide for different expiries: 2Y, 5Y, 10Y, 20Y, 30Y. The calibrated parameters for the SABR model are the same as those in table 7.1 and the SABR-MR parameters are listed in the following table.

β	a	c	κ	ν	ρ
0	0.006	0.065	0.098	0.317	0.033
0.5	0.027	0.068	0.050	0.278	-0.233

Table 7.2: Parameters for the Normal SABR-MR and DD-SABR-MR models for all expiries.

The results for the practical delta and vega are displayed in the following figures.

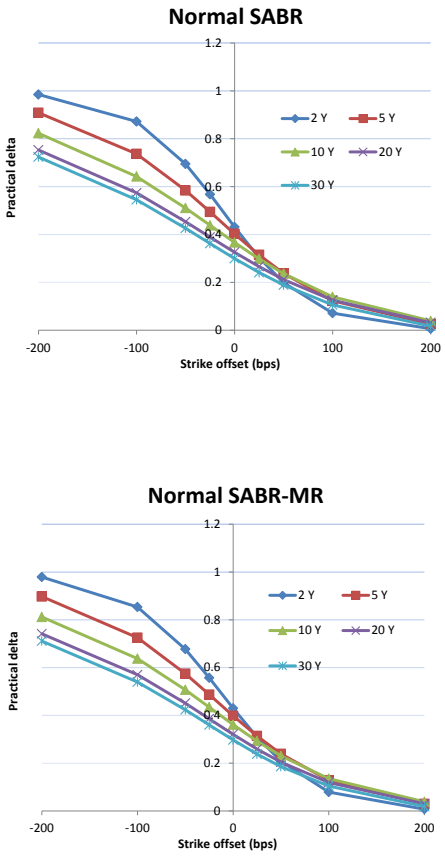


Figure 7.3: Practical delta under different models.

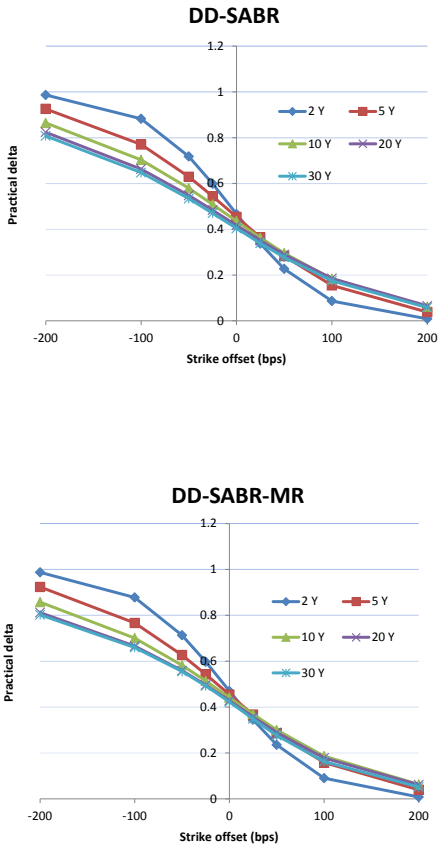


Figure 7.4: Practical delta under different models.

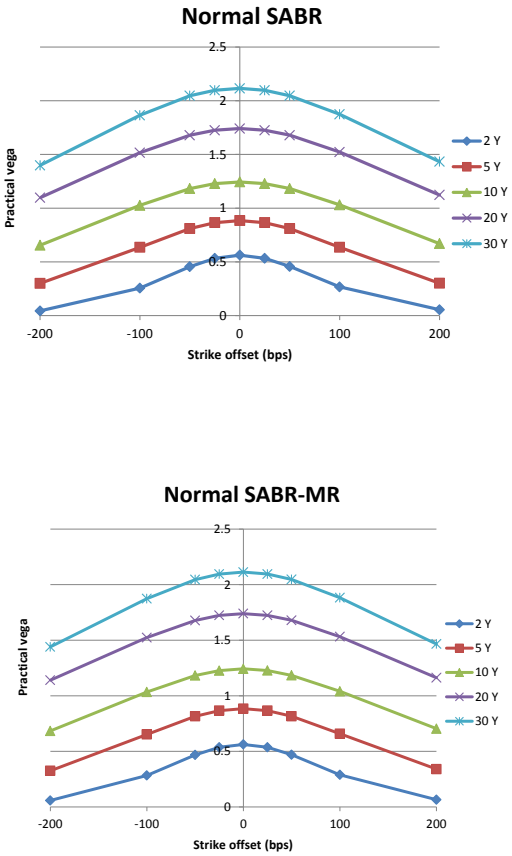


Figure 7.5: Practical vega under different models.

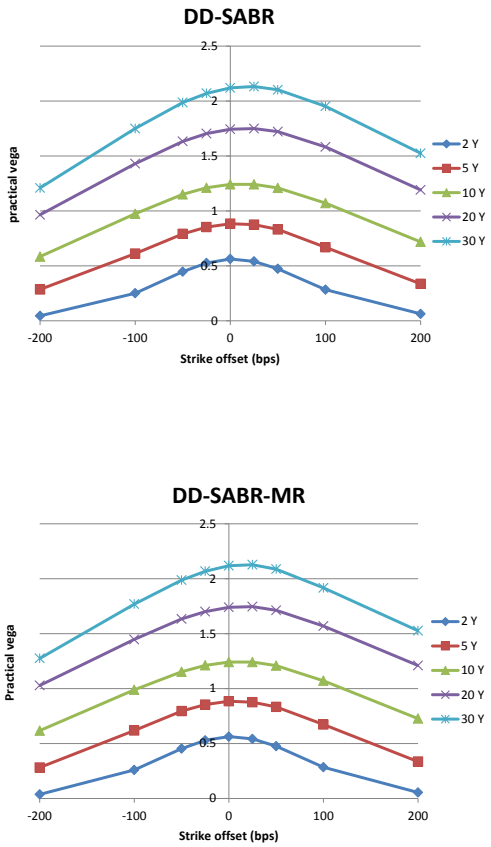


Figure 7.6: Practical vega under different models.

It is seen from the figures that the deltas and vegas computed are almost the same for the (DD-)SABR and (DD-)SABR-MR models across all expiries. We have also tested the second order Greeks and observed a similar conclusion. This is quite surprising since one would expect some key differences in the two models' dynamics to show up here. This concludes that once we properly fit the two models to market smiles, the difference in hedge ratios for the Europeans are not of practical significance.

Chapter 8

Discussion

In this thesis, further developments on Markov-functional models and stochastic volatility models were presented. The main technical contributions were addressed in Chapters 3, 5 and 6.

In Chapter 3 of Part I of the thesis, we drew attention to the problem of pricing and hedging Bermudan swaptions within a one-factor Markov-functional model. Our results showed that different parametrizations of the model could lead to very different hedging results. The innovations on the driving Markov process that we presented in this chapter can potentially prove to be valuable for hedging other interest rate exotics.

Part II (Chapters 4,5,6 and 7) of the thesis contains some important results for several particular aspects of stochastic volatility modelling. We focused on efficient implementations of the SABR model and its extension with mean reversion to price European options and aimed at bringing these numerical schemes to trading desks. In Chapter 5, we worked with the SABR model which has received a lot of attention worldwide. Our results showed that the approximation we derived for the model works extremely well, and could replace the existing schemes in practice which are currently the market standards, e.g. one of the most popular approximations for the SABR model is the SABR formula derived in Hagan et al. [2002]. Chapter 6 is similar in scope but we looked at an extension of the SABR model with mean reversion. An approximation using a similar approach was derived and the results are still quite outstanding although the model has a much more complicated structure.

Chapter 7 is a bit different in focus compared with Chapters 5 and 6. We test further the appropriateness of our chosen models. The hedging properties were assessed for the SABR and SABR-MR models presented in Chapters 5 and 6 in the context of Europeans. We first provided a brief guide and explanation about different ways to hedge a European option with stochastic volatility in theory and practice. We then calculated the deltas and vegas under the two models following

these different methods and the results showed that the SABR and SABR-MR models yield very similar hedging properties for European options.

Bibliography

L. Andersen and V. Piterbarg. *Interest Rate Modeling, 1 edn.* Atlantic Financial Press, 2010.

L. B. G. Andersen and J. Andreasen. Volatile volatilities. *Risk*, 15(2):163–168, 2002.

A. Antonov and M. Spector. Advanced Analytics for the SABR Model. `http://ssrn.com/abstract=2026350`, March 2012.

Ole E. Barndorff-Nielsen. Normal Inverse Gaussian Distributions and Stochastic Volatility Modelling. *Scandinavian Journal of Statistics*, 24(1):1–13, 1997.

B. Bartlett. Hedging under SABR Model. *Wilmott Magazine*, July/August:2–4, 2006.

M. Bennett and J. E. Kennedy. A Comparison of Markov Functional and Market Models: The One Dimensional Case. *The Journal of Derivatives*, 13 (2):22–43, 2005.

H. Berestycki, J. Busca, and I. Florent. Computing the implied volatility in stochastic volatility models. *Comm. Pure Appl. Math.*, 57(10):1352–1373, 2004.

F. Black. The pricing of commodity contracts. *Journal of Financial Economics*, 3(2):167–179, 1976.

F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–659, 1973.

A. Brace, D. Gatarek, and M. Musiela. The market model of interest rate dynamics. *Mathematical Finance*, 7(2):127–155, 1997.

D. Brigo and F. Mercurio. *Interest Rate Models, Theory and Practice.* Springer, 2001.

P. Carr. A survey of preference free option valuation with stochastic volatility. In *Risk's 5th annual European derivatives and risk management congress, Paris*, 2000.

P. Cotton, J.-P. Fouque, G. Papanicolaou, and R. Sircar. Stochastic Volatility Corrections for Interest Rate Derivatives. *Mathematical Finance*, 14(2):173–200, 2004.

J. C. Cox. The constant elasticity of variance option pricing model. *Journal of Portfolio Management*, 23:15–17, 1996.

E. Derman and I. Kani. Riding on a smile. *Risk Magazine*, 7(2):32–39, 1994.

E. Derman and I. Kani. Stochastic implied trees: Arbitrage pricing with stochastic term and strike structure of volatility. *International Journal of Theoretical & Applied Finance*, 1(1):61–110, 1998.

J. Driessen, P. Klassen, and B. Melenberg. The performance of multi-factor term structure models for pricing and hedging caps and swaptions. *Journal of Financial and Quantitative Analysis*, 38(3):635–672, 2003.

B. Duprie. Pricing with a smile. *Risk Magazine*, 7(1):18–20, 1994.

R. Fan, A. Gupta, and P. Ritchken. Hedging in the possible presence of unspanned stochastic volatility: Evidence from swaption markets. *Journal of Finance*, 58(5): 2219–2248, 2003.

J.-P. Fouque, G. Papanicolaou, and R. Sircar. Mean-Reverting Stochastic Volatility. *International Journal of Theoretical & Applied Finance*, 3(1):101–142, 2000a.

J.-P. Fouque, G. Papanicolaou, and R. Sircar. *Derivatives in financial markets with stochastic volatility*. Cambridge University Press, 2000b.

C. Fries. *Mathematical finance. Theory, Modelling, Implementation*. Wiley, 2007.

C. Fries and F. Eckstaedt. A Hybrid Markov-Functional model with simultaneous calibration to the interest rate and FX smile. *Quantitative Finance*, 11(4):587–597, 2009.

C. Fries and M. Rott. Cross-currency and hybrid markov-functional models. *Working paper*, 2004.

D. Gasbarra, T. Sottinen, and E. Valkeila. Gaussian bridges. *Stochastic analysis and applications*, 2:361–382, 2007.

P. Hagan, D. Kumar, A. Lesniewski, and D. Woodward. Managing smile risk. *Wilmott Magazine*, 1(8):84–108, 2002.

P. Hagan, A. Lesniewski, and D. Woodward. Probability distribution in the SABR model of stochastic volatility. *Preprint.*, 2005. Available at http://lesniewski.us/papers/working/ProbDistrForSABR.pdf.

P. Henry-Labordere. A general asymptotic implied volatility for stochastic volatility models. *SSRN eLibrary*, 2005. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract$_id$ = 698601.

S. L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6(2):327–343, 1993.

J. Hull and A. White. The Pricing of Options on Assets with Stochastic Volatilities. *The Journal of Finance*, 42(2):281–300, 1987.

J. Hull and A. White. Pricing interest rate derivatives securities. *The Review of Financial Studies*, 3(4):573–592, 1990.

P. J. Hunt and J. E. Kennedy. *Financial Derivatives in Theory and Practice*. Revised edition, Wiley, 2004.

P. J. Hunt, J. E. Kennedy, and A. Pelsser. Markov functional interest rate models. *Finance and Stochastics*, 4(4):392–408, 2000.

F. Jamshidian. Libor and swap market models and measures. *Finance and Stochastics*, 1(4):293–330, 1997.

S. Johnson and B. Nonas. Arbitrage-free construction of the swaption cube. *Working paper*, 2009.

M. Joshi and R. Rebonato. A displaced diffusion stochastic volatility LIBOR market model: motivation, definition and implementation. *Quantitative Finance*, 3(6):458–469, 2003.

L. Kaisajuntti and J. E. Kennedy. Stochastic Volatility for Interest Rate Derivatives. `http://ssrn.com/abstract=1894887`, September 2011.

L. Kaisajuntti and J. E. Kennedy. An n-Dimensional Markov-functional Interest Rate Model. *Journal of Computational Finance*, Forthcoming.

I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus, 2nd edition*. Springer-Verlag, Berlin, Heidelberg and New York, 1991.

J. E. Kennedy and D. Pham. Implications for Hedging of the choice of driving process for one-factor Markov-functional models, December 2011. Available at SSRN: http://ssrn.com/abstract=1972932 or http://dx.doi.org/10.2139/ssrn.1972932.

J. E. Kennedy, S. Mitra, and D. Pham. On the approximation of the SABR model: a probabilistic approach. *Applied Mathematical Finance*, Forthcoming.

K. Larsson. Dynamic Extensions and Probabilistic Expansions of the SABR model. http://ssrn.com/paper=1536471, 2010.

F. A. Longstaff and E. S. Schwartz. Valuing American options by simulation: A simple least-squares approach. *Review of Financial Studies*, 14(1):113–147, 2001.

F. A. Longstaff, P. Santa-Clara, and E. S. Schwartz. Throwing away a billion dollars: The cost of suboptimal exercise strategies in the swaptions market. *Journal of Financial Economics*, 62(1):39–66, 2001.

Y. Maghsoodi. Exact solution of a martingale stochastic volatility option problem and its empirical evaluation. *Mathematical Finance*, 17(2):249–265, April 2007.

D. Marris. Financial option pricing and skewed volatility. Master's thesis, University of Cambridge, 1999.

J. Masoliver and J. Perello. Multiple time scales and the exponential Ornstein-Uhlenbeck stochastic volatility model. *Quantitative Finance*, 6(5):423–433, 2006.

F. Mercurio and M. Morini. A Note on Hedging with Local and Stochastic Volatility Models. http://ssrn.com/abstract=1294284 or http://dx.doi.org/10.2139/ssrn.1294284, November 3 2008.

K. R. Miltersen, K. Sandmann, and D. Sondermann. Closed form solutions for term structure derivatives with log-normal interest rates. *Journal of Finance*, 52(1):409–430, 1997.

S. Mitra. A One Factor Approximation for the SABR model. Master's thesis, Warwick Business School, University of Warwick, 2010.

M. Musiela and M. Rutkowski. *Martingale Methods in Financial Modelling*. Second edition, Springer-Verlag, 2004.

J. Obloj. Fine-tune your smile: Correction to hagan et al. *Wilmott Magazine*, 35: 102–104, 2008.

L. Paulot. Asymptotic Implied Volatility at the Second Order with Application to the SABR model. http://ssrn.com/abstract=1413649, 2009.

A. Pelsser. *Efficient Methods for Valuing Interest Rate Derivatives*. Springer, 2000.

J. Perello, R. Sircar, and J. Masoliver. Option pricing under stochastic volatility: the exponential Ornstein-Uhlenbeck model. *Journal of Statistical Mechanics*, P06010, 2008.

I. O. Pertursson. Comparison of the hedging performance of Bermudan swaptions for different driving processes in the Markov functional interest rate model. Master's thesis, University of Warwick, 2008.

R. Pietersz. *Pricing Models for Bermudan-Style Interest Rate Derivatives*. PhD thesis, Erasmus University Rotterdam, Rotterdam, The Netherlands, December 2005. URL `http://hdl.handle.net/1765/7122`.

R. Pietersz and A. Pelsser. Risk Managing Bermudan Swaptions in the Libor BGM Model. *The Journal of Derivatives*, 11(3):51–62, 2004.

R. Pietersz and A. Pelsser. A comparison of single factor Markov-functional and multi factor market models. *Reviews of Derivatives Research*, 13 (3):245–272, 2010.

R. Pietersz, A. Pelsser, and M.V. Regenmortel. Fast drift approximated pricing in the bgm model. *Journal of Computational Finance*, 8(1):93–124, 2004.

V. Piterbarg. Stochastic volatility model with time-dependent skew. *Applied Mathematical Finance*, 12(2):147–185, 2005.

V. V. Piterbarg. TARNs: Models, valuation, risk sensitivities. *Wilmott Magazine*, 14: 62–71, 2004.

R. Rebonato. *Modern Pricing of Interest-Rate Derivatives*. Princeton University Press, 2002.

R. Rebonato. *Volatility and Correlation, Second Edition*. Wiley, 2004.

R. Rebonato, K. Mckay, and R. White. *The SABR/LIBOR Market Model: Pricing, Calibration and Hedging for Complex Interest-Rate Derivatives*. Wiley, 2009.

M. Rubinstein. Displaced diffusion option pricing. *Journal of Finance*, 38(1):213–217, 1983.

M. Schroder. Computing the constant elasticity of variance option pricing formula. *Journal of Finance*, 44(1):211–219, 1989.

L. Scott. Option Pricing when the Variance Changes Randomly: Theory, Estimation and an Application. *Journal of Financial and Quantitative Analysis*, 22(4):419–438, 1987.

S. Svoboda-Greenwood. Displaced diffusion as an approximation of the constant elasticity of variance. *Applied Mathematical Finance*, 16(3):269–286, 2009.

Q. Wu. Series Expansion of the SABR Joint Density. *Mathematical Finance*, 2010. Available at http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9965.2010.00460.x/abstract.