On the approximation of the SABR model: a probabilistic approach

Joanne E. Kennedy, Subhankar Mitra, and Duy Pham
Department of Statistics, University of Warwick

April 21, 2012

Abstract

In this paper, 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 elastic 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 the 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 Volvol.

Key words: the SABR model, displaced diffusion, stochastic volatility .

Contents

1 Introduction

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&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. 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 elastic 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 & Rebonato (2003), Piterbarg (2005) and Larsson (2010). In this paper, 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 paper is organized as follows. Section 2 compares the SABR model and its displaced diffusion version with the mapping connecting them numerically. We develop our approximation in section 3 where we quote the appropriate formulae and match the parameters for implementation. In section 4, we numerically investigate the quality of our approximation in conjunction with other approximations and Monte Carlo simulations. Section 5 concludes the paper.

2 SABR model

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

$$\begin{aligned} dF_t &= \sigma_t F_t^{\beta} dW_t \qquad \beta \in [0,1], \\ d\sigma_t &= \nu \sigma_t dZ_t \qquad \nu > 0, \end{aligned} \tag{2.0.1}$$

where W_t and Z_t are correlated Brownian motions such that $dW_t dZ_t = \rho dt$ 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. Hagan et al. (2002) was the first to point out these roles through their SABR formula which will be introduced in section 3.2.2. The parameter β has a primary effect on the skew, i.e.

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

reducing β from 1 to 0 gives rise 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.

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

$$\begin{aligned} dF_t &= \hat{\sigma}_t(F_t + \theta)dW_t, \\ d\hat{\sigma}_t &= \nu\hat{\sigma}_t dZ_t, \\ dW_t dZ_t &= \rho dt. \end{aligned} \tag{2.1.1}$$

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 (2.0.1) and (2.1.1) 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 paper, 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.

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.

Figure 2.1 illustrates 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 2.2). 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.

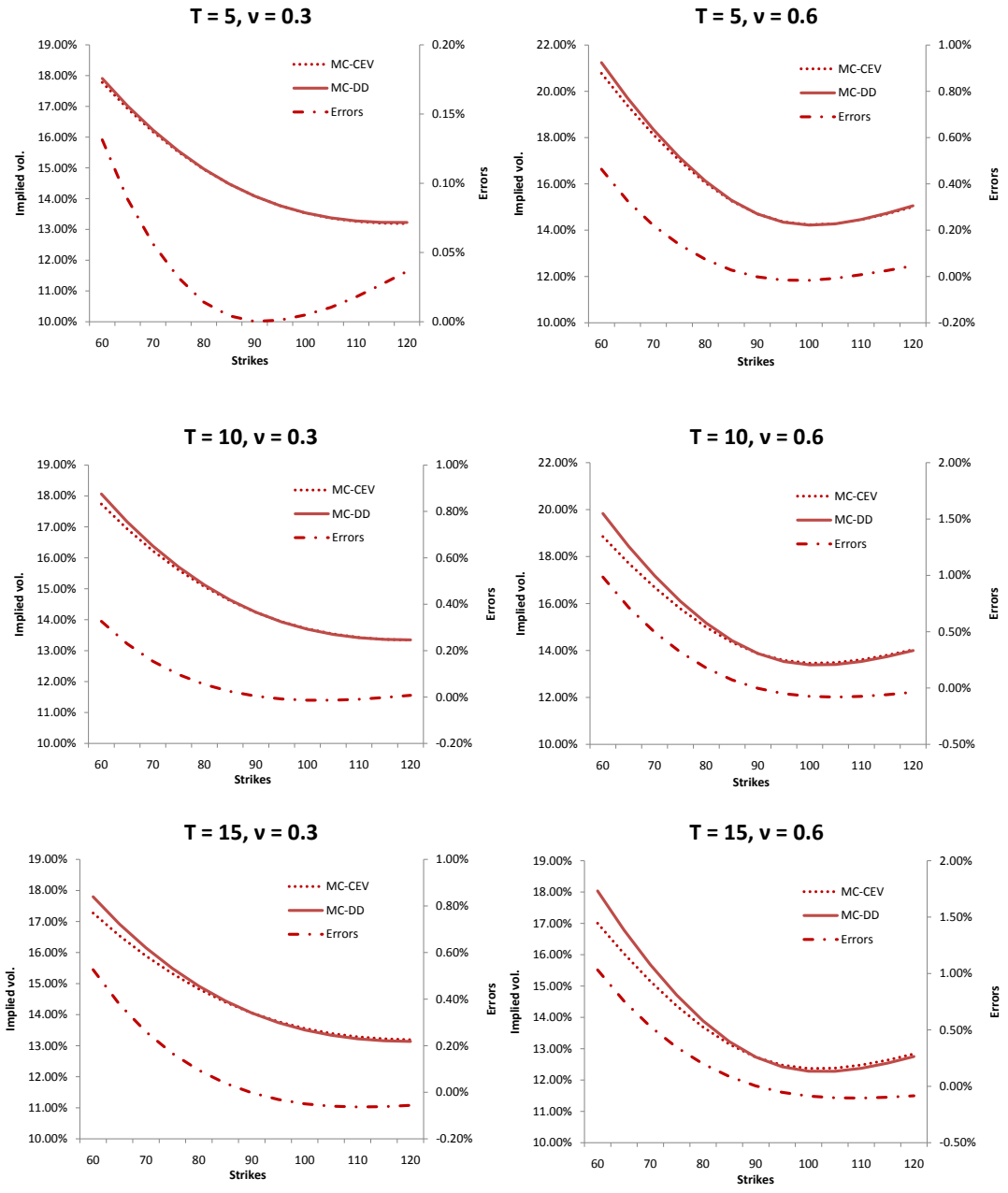

Figure 2.1: Effects of maturity T and Volvol ν on the mapping when the ATM are matched. Parameters: $\beta = 0.5, \rho = -0.2, \sigma_0 = 130\%, F_0 = 90$. MC-CEV: CEV-SABR MC solution, MC-DD: DD-SABR MC solution, Errors: MC-DD minus MC-CEV.

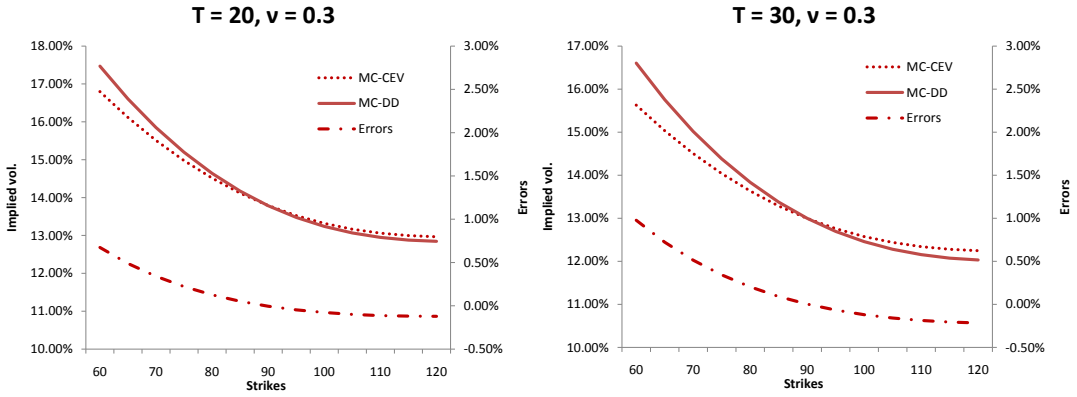

Figure 2.2: Effect of very long maturity T on the mapping when the ATM are matched. Parameters: $\beta = 0.5, \rho = -0.2, \sigma_0 = 130\%, F_0 = 90$.

We mentioned that both ρ and β affect the skew. In figure 2.3, 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 2.4 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.

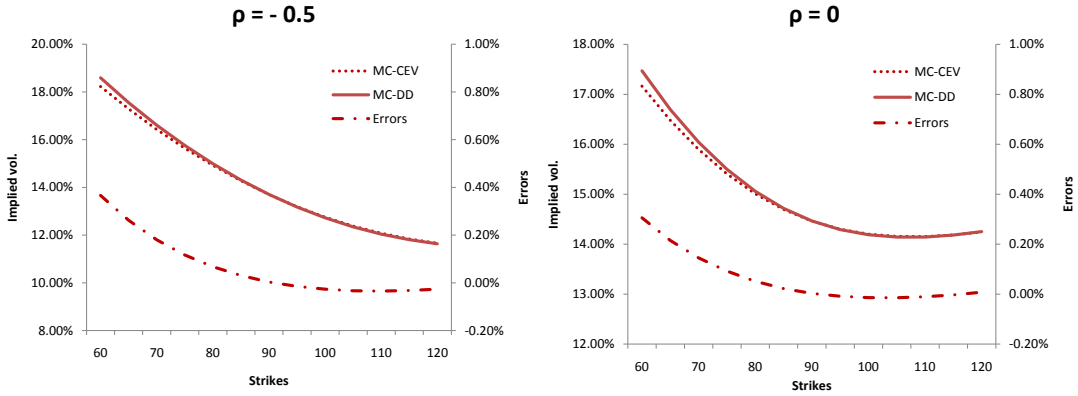

Figure 2.3: Effect of ρ on the mapping when the ATM are matched. Parameters: $\beta = 0.5, T = 10, \nu = 0.3, \sigma_0 = 130\%, F_0 = 90$.

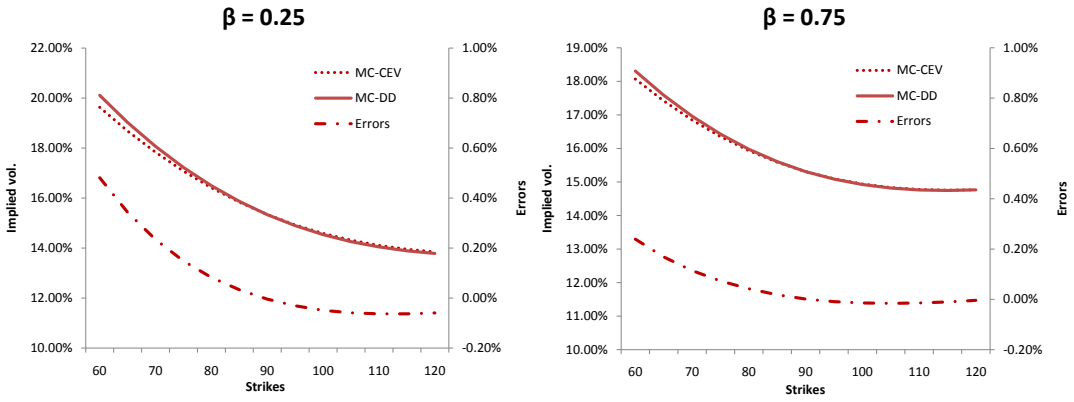

Figure 2.4: Effect of β on the mapping when the ATM are matched. Parameters: $\rho = -0.2, T = 10, \nu = 0.3, F_0 = 90$, σ_0 is chosen for each case so that the ATM are comparable.

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 (figure 2.5). 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 figure 2.5 in table 2.1. For the 20 year maturity case, it is seen that around a quarter of the mass is given to the absorbing barrier and a fifth to 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 useful for all different asset classes including interest rates.

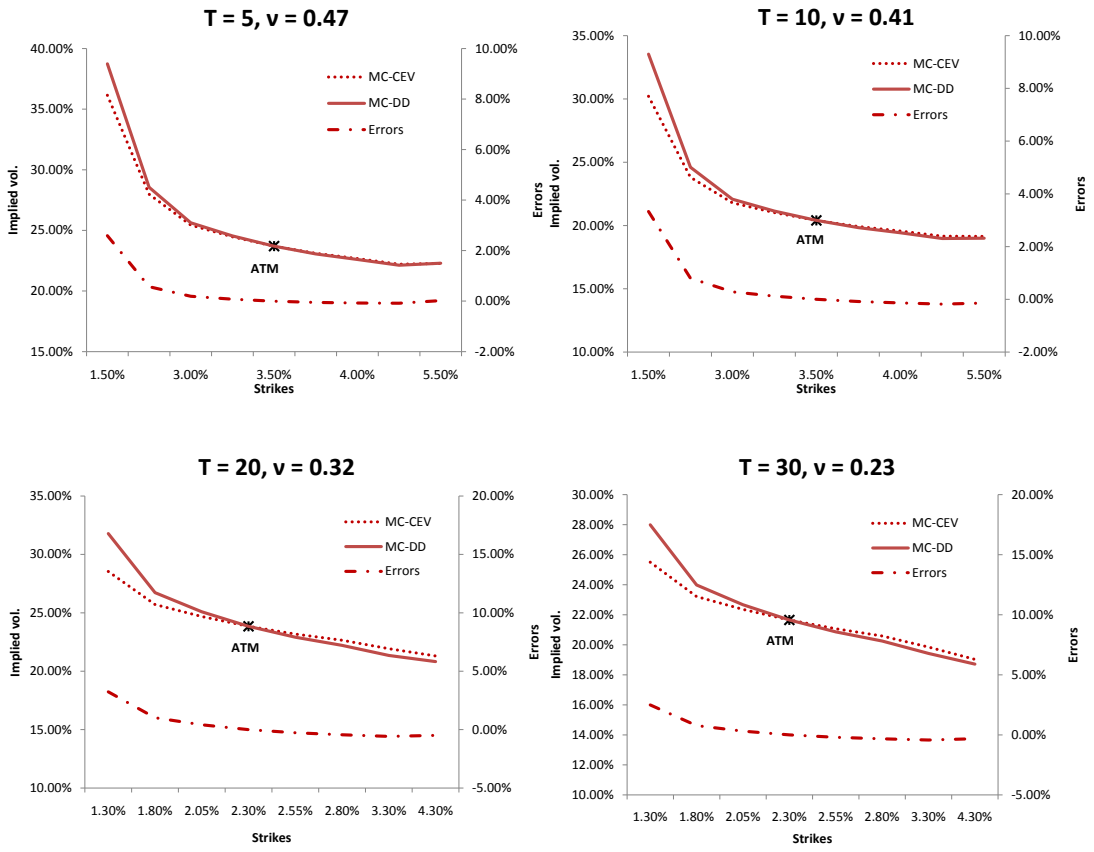

Figure 2.5: Implied volatilities under different models. Parameters: $\beta = 0.5, \rho = -0.2, \sigma_0 = 4.30\%, 3.80\%, 4.10\%, 3.70\%$ as maturity increases respectively.

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 2.1: Probability mass assigned to the absorbing barrier (CEV-SABR) and the negative rates region (DD-SABR) for the four cases considered in figure 2.5 (computed by direct Monte Carlo simulation).

3 A probabilistic approximation

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

$$\begin{aligned} \mathbb{C}_0(K, F_0) &= \mathbb{E}\left[(F_T - K)^+\right] \\ &= \mathbb{E}\left[\mathbb{E}\{(F_T - K)^+ | \sigma_T\}\right]. \end{aligned} \tag{3.1.1}$$

Assuming that we have in mind some distribution for $F_T|\sigma_T$: the conditional distribution of F_T given σ_T (see sections 3.2 and 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 A for details).

- **Normal SABR**:

$$\begin{aligned} s_T &\triangleq F_0 + \frac{\rho}{\nu}(\sigma_T - \sigma_0) + \sqrt{1-\rho^2} V_T^{\frac{1}{2}} G, \\ g(s) &= s. \end{aligned} \tag{3.1.2}$$

- **Log-Normal SABR**:

$$\begin{aligned} s_T &\triangleq \ln F_0 + \frac{\rho}{\nu}(\sigma_T - \sigma_0) - \frac{1}{2} V_T + \sqrt{1-\rho^2} V_T^{\frac{1}{2}} G, \\ g(s) &= e^s. \end{aligned} \tag{3.1.3}$$

- **DD-SABR**:

$$\begin{aligned} s_T &\triangleq \ln(F_0 + \theta) + \beta F_0^{\beta-1} \frac{\rho}{\nu}(\sigma_T - \sigma_0) - \frac{1}{2}\beta^2 F_0^{2\beta-2} V_T + \beta F_0^{\beta-1}\sqrt{1-\rho^2} V_T^{\frac{1}{2}} G, \\ \theta &= F_0 \frac{1-\beta}{\beta}, \\ g(s) &= e^s - \theta. \end{aligned} \tag{3.1.4}$$

Here $V_T := \int_0^T \sigma_t^2 dt$ 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. $d\sigma_t = \nu \sigma_t dZ_t$ where Z is a Brownian motion. The first two* ***conditional moments of the realized variance*** *V_T have the following analytical expressions:*

$$\mathbb{E}(V_T|\sigma_T) = \frac{\sigma_0^2\sqrt{T}}{2\nu} \frac{\left[\Phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}} + \nu\sqrt{T}\right) - \Phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}} - \nu\sqrt{T}\right)\right]}{\phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}} + \nu\sqrt{T}\right)}, \tag{3.1.5}$$

$$\begin{aligned} \mathbb{E}(V_T^2|\sigma_T) &= -\frac{\sigma_0^4\sqrt{T}}{4\nu^3}\left(1 + e^{2\ln(\sigma_T/\sigma_0)}\right) \frac{\left[\Phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}} + \nu\sqrt{T}\right) - \Phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}} - \nu\sqrt{T}\right)\right]}{\phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}} + \nu\sqrt{T}\right)} \\ &\quad + \frac{\sigma_0^4\sqrt{T}}{4\nu^3} \frac{\left[\Phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}} + 2\nu\sqrt{T}\right) - \Phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}} - 2\nu\sqrt{T}\right)\right]}{\phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}} + 2\nu\sqrt{T}\right)}, \end{aligned} \tag{3.1.6}$$

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

Proof: Appendix B.

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 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 (3.1.1) 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**:*

$$\begin{aligned}\mu(\sigma_T) &= F_0 + \frac{\rho}{\nu}(\sigma_T - \sigma_0),\\ \eta^2(\sigma_T) &= (1-\rho^2)\frac{\sigma_0^2\sqrt{T}}{2\nu}\frac{\left[\Phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}} + \nu\sqrt{T}\right) - \Phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}} - \nu\sqrt{T}\right)\right]}{\phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}} + \nu\sqrt{T}\right)}.\end{aligned} \tag{3.2.1}$$

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

$$\begin{aligned}\mu(\hat{\sigma}_T) &= \ln(F_0+\theta) + \frac{\rho}{\nu}(\hat{\sigma}_T - \hat{\sigma}_0) - \frac{\hat{\sigma}_0^2\sqrt{T}}{4\nu}\frac{\left[\Phi\left(\frac{\ln(\hat{\sigma}_T/\hat{\sigma}_0)}{\nu\sqrt{T}} + \nu\sqrt{T}\right) - \Phi\left(\frac{\ln(\hat{\sigma}_T/\hat{\sigma}_0)}{\nu\sqrt{T}} - \nu\sqrt{T}\right)\right]}{\phi\left(\frac{\ln(\hat{\sigma}_T/\hat{\sigma}_0)}{\nu\sqrt{T}} + \nu\sqrt{T}\right)},\\ \eta^2(\hat{\sigma}_T) &= \frac{\hat{\sigma}_0^2\sqrt{T}}{2\nu}\left((1-\rho^2) - \frac{(\hat{\sigma}_T^2+\hat{\sigma}_0^2)}{8\nu^2}\right)\frac{\left[\Phi\left(\frac{\ln(\hat{\sigma}_T/\hat{\sigma}_0)}{\nu\sqrt{T}} + \nu\sqrt{T}\right) - \Phi\left(\frac{\ln(\hat{\sigma}_T/\hat{\sigma}_0)}{\nu\sqrt{T}} - \nu\sqrt{T}\right)\right]}{\phi\left(\frac{\ln(\hat{\sigma}_T/\hat{\sigma}_0)}{\nu\sqrt{T}} + \nu\sqrt{T}\right)}\\ &\quad + \frac{\hat{\sigma}_0^4\sqrt{T}}{16\nu^3}\frac{\left[\Phi\left(\frac{\ln(\hat{\sigma}_T/\hat{\sigma}_0)}{\nu\sqrt{T}} + 2\nu\sqrt{T}\right) - \Phi\left(\frac{\ln(\hat{\sigma}_T/\hat{\sigma}_0)}{\nu\sqrt{T}} - 2\nu\sqrt{T}\right)\right]}{\phi\left(\frac{\ln(\hat{\sigma}_T/\hat{\sigma}_0)}{\nu\sqrt{T}} + 2\nu\sqrt{T}\right)}\\ &\quad - \frac{\hat{\sigma}_0^4 T}{16\nu^2}\left(\frac{\left[\Phi\left(\frac{\ln(\hat{\sigma}_T/\hat{\sigma}_0)}{\nu\sqrt{T}} + \nu\sqrt{T}\right) - \Phi\left(\frac{\ln(\hat{\sigma}_T/\hat{\sigma}_0)}{\nu\sqrt{T}} - \nu\sqrt{T}\right)\right]}{\phi\left(\frac{\ln(\hat{\sigma}_T/\hat{\sigma}_0)}{\nu\sqrt{T}} + \nu\sqrt{T}\right)}\right)^2,\end{aligned} \tag{3.2.2}$$

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 C for more details). Clearly, the formulae in (3.2.2) for the Log-Normal SABR model is obtained when $\beta = 1$.

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 (3.1.1) has the equivalent expression $\mathbb{E}\left[(F_T-K)^+|Z_T\right]$ and (3.1.1) now reads

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

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**:

$$\mathbb{C}_0(K,F_0) = \int_{-\infty}^{\infty} \left[\sqrt{\eta^2(x)}\phi\left(\frac{K-\mu(x)}{\sqrt{\eta^2(x)}}\right) + (\mu(x)-K)\left(1-\Phi\left(\frac{K-\mu(x)}{\sqrt{\eta^2(x)}}\right)\right)\right] \frac{e^{\frac{-x^2}{2T}}}{\sqrt{2\pi T}}dx. \tag{3.2.3}$$

2. **Log-Normal SABR**:

$$\mathbb{C}_0(K,F_0) = \int_{-\infty}^{\infty} \left[e^{\mu(x)+\frac{\eta^2(x)}{2}}\Phi\left(\frac{\mu(x)+\eta^2(x)-\ln K}{\sqrt{\eta^2(x)}}\right) - K\Phi\left(\frac{\mu(x)-\ln K}{\sqrt{\eta^2(x)}}\right)\right] \frac{e^{\frac{-x^2}{2T}}}{\sqrt{2\pi T}}dx. \tag{3.2.4}$$

Remark 1 *: For the DD-SABR model, we have exactly the same formula as* (3.2.4) *with K replaced by $K+\theta$.*

Both (3.2.3) and (3.2.4) 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 2 *: 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* (3.1.6)*. For certain parameter choices, the growth rate of $\mathbb{E}[(g(s_T)-K)^+|Z_T=x]$ in equation* (3.2.4) *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*

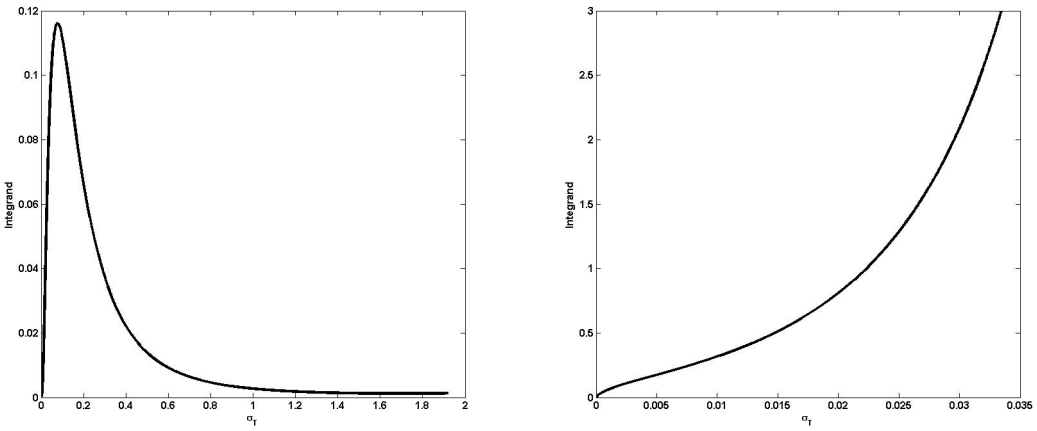

Figure 3.1: The integrand of (3.2.4) as a function of σ_T. Left plot: $\beta = 1, \rho = -0.5, F_0 = 90, K = 90, T = 10, \nu = 0.3, \sigma_0 = 15\%$, right plot: $\beta = 1, \rho = -0.5, F_0 = 90, K = 90, T = 15, \nu = 0.6, \sigma_0 = 15\%$.

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

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

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

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

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 3.1 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 et al. (2009)	not available	tested	not tested	not available
Wu (2010)	tested	tested	tested	not available
Larsson (2010)	tested	tested	not available	tested

Table 3.1: Check list 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 $\sigma_B(K, S_0)$ for a Vanilla call (or put) option written on the forward price S struck at some strike K has the following form

$$\begin{aligned}\sigma_B(K,F_0) &= \frac{\sigma_0}{(F_0K)^{(1-\beta)/2}\left\{1+\frac{(1-\beta)^2}{24}\ln^2\frac{F_0}{K}+\frac{(1-\beta)^4}{1920}\ln^4\frac{F_0}{K}+...\right\}}\left(\frac{z}{x(z)}\right)\\ &\quad\left\{1+\left[\frac{(1-\beta)^2}{24}\frac{\sigma_0^2}{(F_0K)^{1-\beta}}+\frac{1}{4}\frac{\rho\beta\sigma_0\nu}{(F_0K)^{\frac{1-\beta}{2}}}+\frac{2-3\rho^2}{24}\nu^2\right]T+...\right\}, \quad (3.2.5)\end{aligned}$$

where

$$\begin{aligned}z &= \frac{\nu}{\sigma_0}(F_0K)^{(1-\beta)/2}\ln\frac{F_0}{K},\\ x(z) &= \ln\left\{\frac{\sqrt{1-2\rho z+z^2}+z-\rho}{1-\rho}\right\}. \quad (3.2.6)\end{aligned}$$

The ATM Black implied volatility reduces to

$$\sigma_B(F_0,F_0)=\sigma_0F_0^{\beta-1}\left\{1+\left[\frac{(1-\beta)^2}{24}\frac{\sigma_0^2}{F_0^{2-2\beta}}+\frac{1}{4}\frac{\rho\beta\sigma_0\nu}{F_0^{1-\beta}}+\frac{2-3\rho^2}{24}\nu^2\right]T+...\right\}. \quad (3.2.7)$$

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

$$\begin{aligned}\sigma_B(K,F_0) &= \hat{\sigma}_0\frac{\sqrt{(F_0+\theta)(K+\theta)}}{\sqrt{F_0K}}\left(\frac{1+\frac{1}{24}\ln^2\frac{F_0+\theta}{K+\theta}+\frac{1}{1920}\ln^4\frac{F_0+\theta}{K+\theta}+\dots}{1+\frac{1}{24}\ln^2\frac{F_0}{K}+\frac{1}{1920}\ln^4\frac{F_0}{K}+\dots}\right)\left(\frac{z}{x(z)}\right)\\ &\quad\left\{1+\left[\frac{2\theta/\sqrt{F_0K}+\theta^2/(F_0K)}{24}\hat{\sigma}_0^2+\frac{1}{4}\rho\nu\hat{\sigma}_0+\frac{2-3\rho^2}{24}\nu^2\right]T+\dots\right\}, \quad (3.2.8)\end{aligned}$$

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

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 (3.2.6). For the special case of the ATM option, the formula reduces to

$$\begin{aligned} \sigma_B(F_0,F_0) &= \hat{\sigma}_0\frac{F_0+\theta}{F_0}\left\{1+\left[\frac{2\theta/F_0+\theta^2/F_0^2}{24}\hat{\sigma}_0^2+\frac{1}{4}\rho\nu\hat{\sigma}_0+\frac{2-3\rho^2}{24}\nu^2\right]T+\ldots\right\} \\ &= \sigma_0F_0^{\beta-1}\left\{1+\left[\frac{2\frac{1-\beta}{\beta}+\frac{(1-\beta)^2}{\beta^2}}{24}\sigma_0^2\beta^2F_0^{2\beta-2}+\frac{1}{4}\rho\nu\sigma_0\beta F_0^{\beta-1}+\frac{2-3\rho^2}{24}\nu^2\right]T+\ldots\right\} \\ &= \sigma_0F_0^{\beta-1}\left\{1+\left[\frac{1-\beta^2}{24}\frac{\sigma_0^2}{F_0^{2-2\beta}}+\frac{1}{4}\frac{\rho\beta\sigma_0\nu}{F_0^{1-\beta}}+\frac{2-3\rho^2}{24}\nu^2\right]T+\ldots\right\}. \end{aligned} \tag{3.2.9}$$

For the rest of the paper, we will refer to (3.2.8) and (3.2.9) as the DD-SABR formula. 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 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 3.1 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 $\nu^2T > 1.6$. The reason is that most of the techniques (singular perturbation or heat kernel expansion) are based on the assumption of small total volatility ν^2T[‡] to allow for accurate asymptotic expansions up to the second order. As discussed in section 3.2.1, the total volatility ν^2T also affects the Normal approximation to some extent. An intuitive reason for this adverse effect is that a larger value of ν^2T will push the true conditional distribution of $s_T|\sigma_T$ further away from Normal. However, in the results presented in section 4, the Normal approximation is shown to perform quite well for the Normal SABR model up to 30 years maturity or very large $\nu^2T \approx 10.8$. For the other sub-models, it works well up to 15 years maturity or $\nu^2T \approx 1.8$. 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 et al. (2009) but the problem still remains.

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

[‡]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).

sample of $s_T|\sigma_T$ against the standard Normal distribution. In figure 3.2, 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.

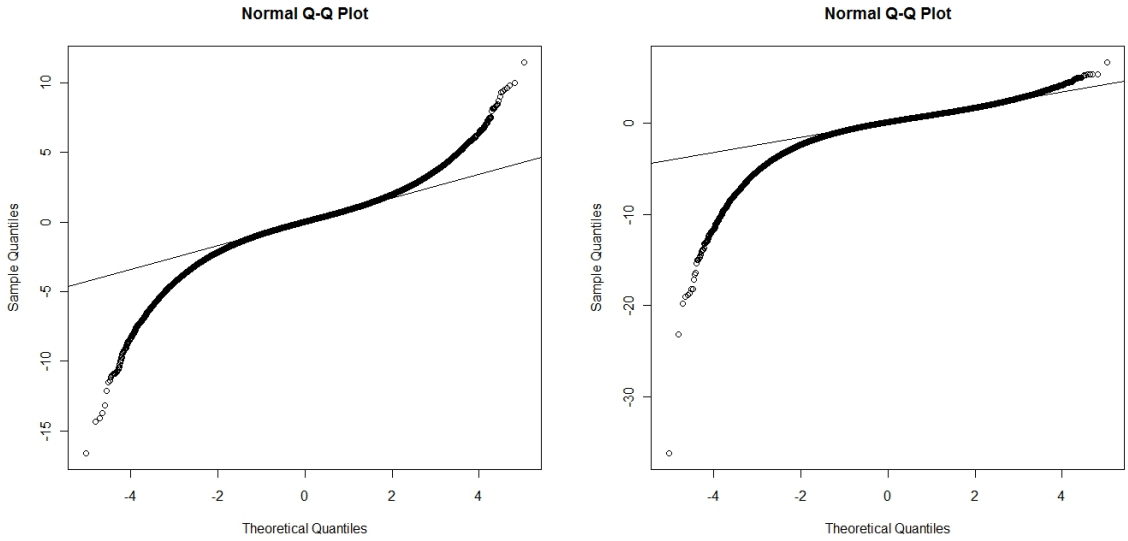

Figure 3.2: Normal Q-Q plots: standardized conditional samples of $s_T|\sigma_T$ against the standard Normal distribution. Common parameters: $\beta = 1, \rho = -0.5, F_0 = 90, \sigma_0 = 5\%$. left plot: $T = 15, \nu = 0.3, \sigma_T = 5\%$, right plot: $T = 15, \nu = 0.3, \sigma_T = 50\%$.

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:

$$f_{NIG}(s; \hat{\alpha}, \hat{\beta}, \hat{\mu}, \hat{\delta}) = \frac{\hat{\alpha}}{\hat{\delta}} \exp(\hat{\delta}\sqrt{\hat{\alpha}^2 - \hat{\beta}^2} - \hat{\beta}\hat{\mu}) \frac{K_1\left(\hat{\alpha}\hat{\delta}\sqrt{1 + (\frac{s-\hat{\mu}}{\hat{\delta}})^2}\right)}{\sqrt{1 + (\frac{s-\hat{\mu}}{\hat{\delta}})^2}} \exp(\hat{\beta}s), \tag{3.3.1}$$

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.

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:

$$X = \hat{\mu} + \hat{\beta}Y + \sqrt{Y}G, \tag{3.3.2}$$

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 (3.1.2) and (3.1.4) that conditioned on σ_T, s_T will have a similar form as (3.3.2). 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.

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

$$\begin{aligned}
\mathbb{C}_0(K, F_0) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(g(s)-K)^+ f_{NIG}(s;\hat{\alpha}(x),\hat{\beta}(x),\hat{\mu}(x),\hat{\delta}(x))ds\frac{e^{-\frac{x^2}{2T}}}{\sqrt{2\pi T}}dx \\
&= \int_{-\infty}^{\infty}\int_{g^{-1}(K)}^{\infty}(g(s)-K) f_{NIG}(s;\hat{\alpha}(x),\hat{\beta}(x),\hat{\mu}(x),\hat{\delta}(x))ds\frac{e^{-\frac{x^2}{2T}}}{\sqrt{2\pi T}}dx, \quad (3.3.3)
\end{aligned}$$

where $g(.)$ (specified in (3.1.2), (3.1.3) and (3.1.4)) 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 (3.3.3) has the following form

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

where f_{NIG} is given by (3.3.1). 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

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

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

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

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 $\int_{l^*}^{\infty} y^n K_1(y)dy$.
- 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. $l_{i-1} \le \tilde{l}(x,K) < l_i$

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

  where $\int_{\tilde{l}(x,K)}^{l_i} y^n K_1(y)dy$ 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.

4 Numerical study

In this section, we investigate the quality of the approximations developed in this paper. 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.

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 figure 4.1

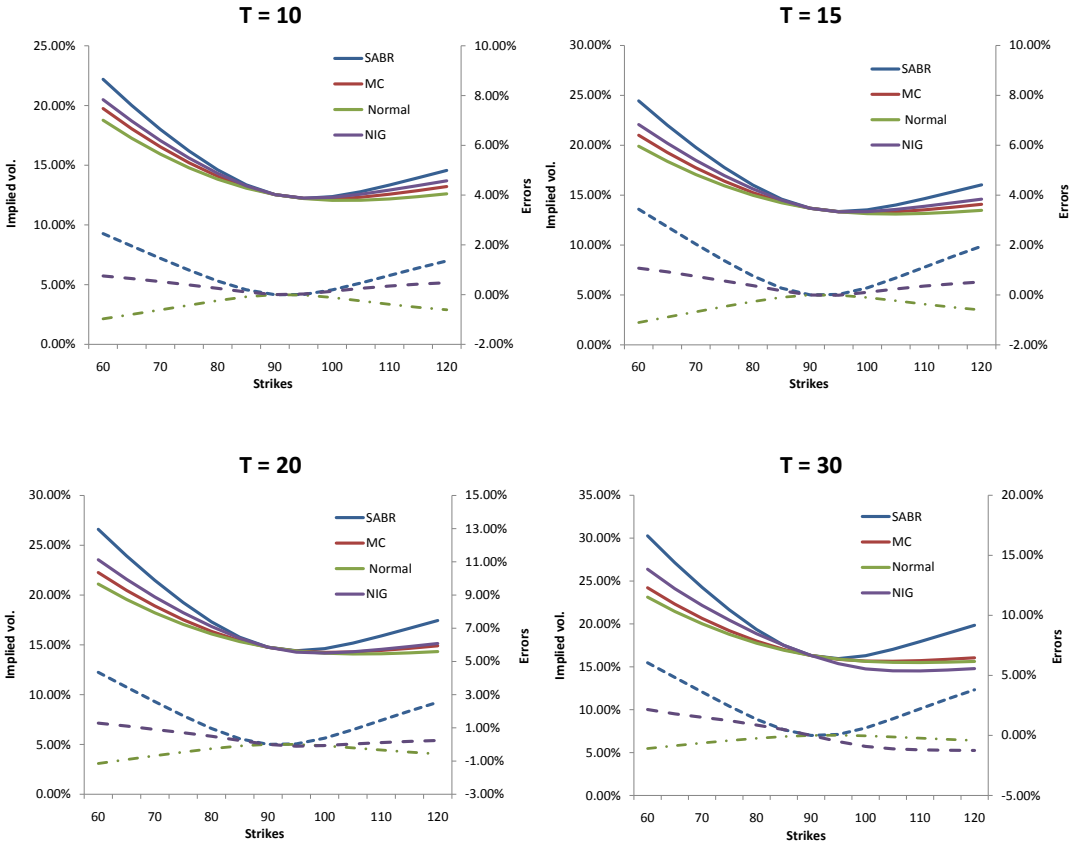

Figure 4.1: Effects of maturity within a high Volvol regime on the Normal and NIG approximations. Other parameters: $\beta = 0, \rho = -0.1, F_0 = 90, \nu = 0.6, \sigma_0 = 9$. The dashed curve of the same colour indicates the errors of the corresponding approximation.

		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 4.1: 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 4.1). 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.

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:

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

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

Figure 4.2 displays the moderate maturity cases where the Normal approximation still performs quite well.

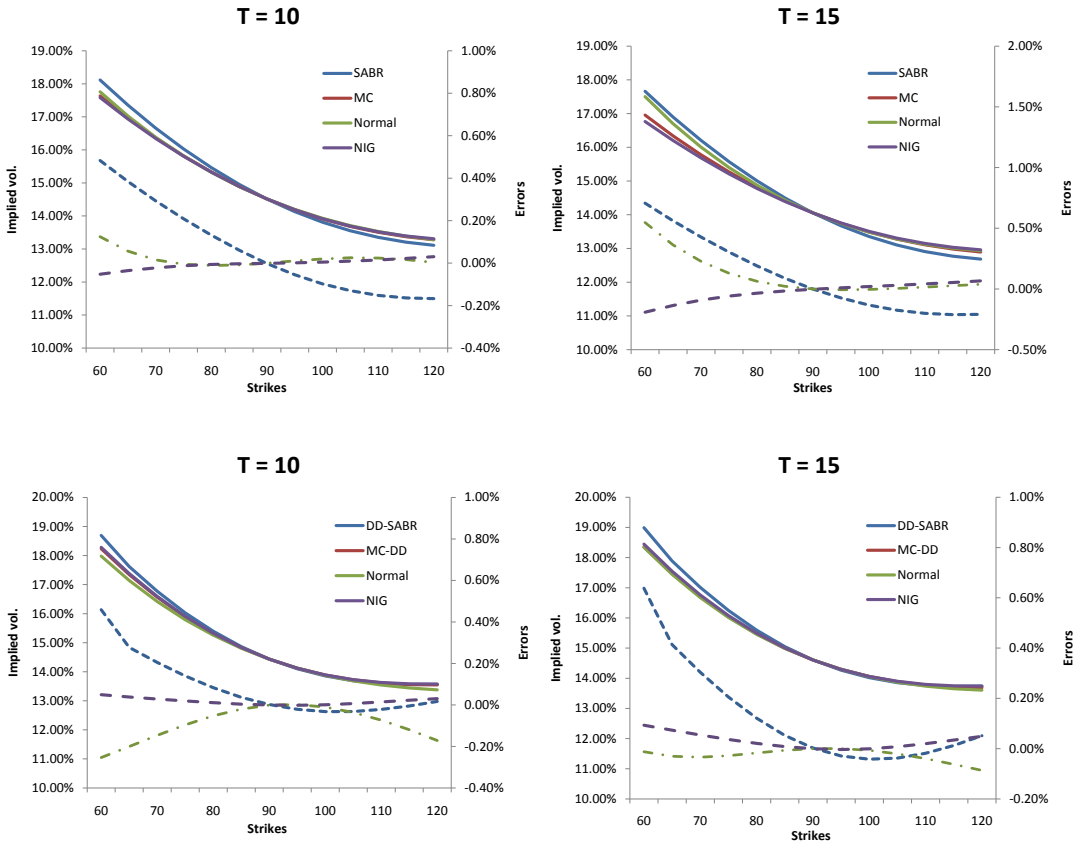

Figure 4.2: Effects of moderate maturity within a low Volvol regime on the Normal and NIG approxmations. Common parameters: $\nu = 0.3, F_0 = 90$, top: $\beta = 1, \sigma_0 = 15\%, \rho = -0.5$, bottom $\beta = 0.5, \sigma_0 = 130\%, \rho = -0.2$. The dashed curve of the same colour indicates the errors 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 2. Although we apply the truncation method mentioned in remark 2, 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 4.3 and 4.4.

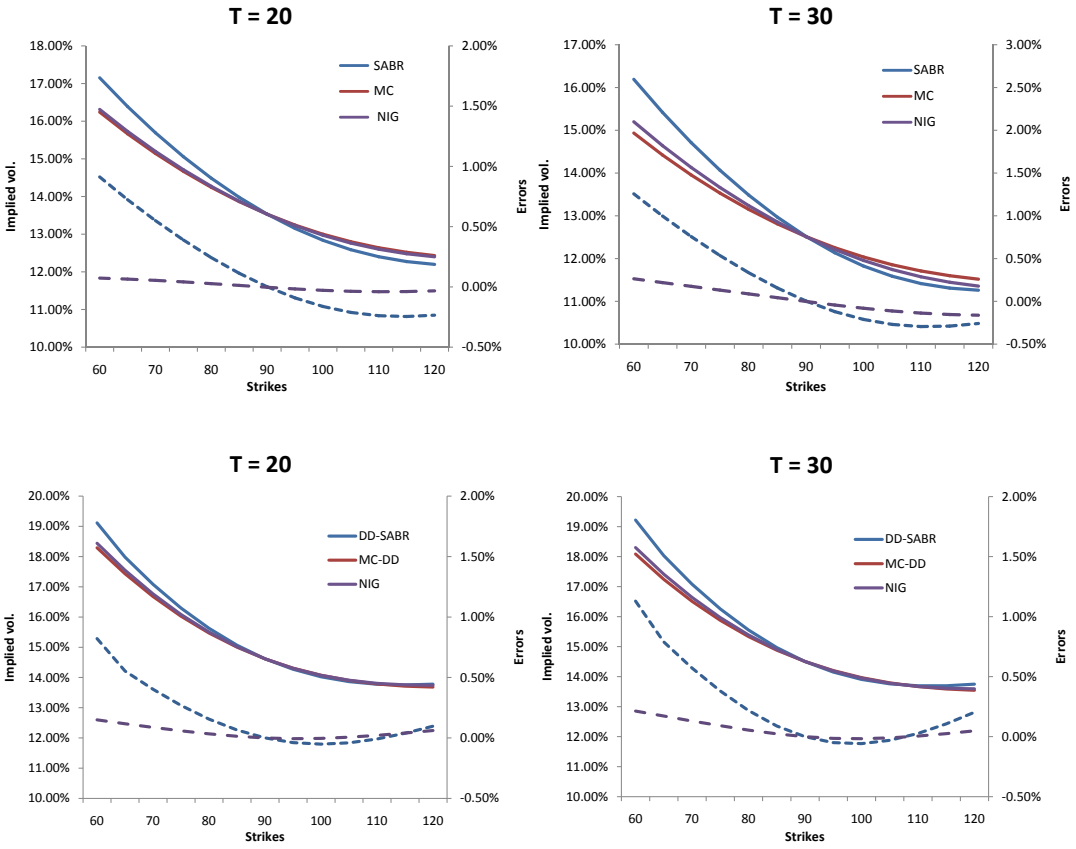

Figure 4.3: Effects of very long maturity within a low Volvol regime on the NIG approximation. Common parameters: $\nu = 0.3, F_0 = 90$, top: $\beta = 1, \sigma_0 = 15\%, \rho = -0.5$, bottom: $\beta = 0.5, \sigma_0 = 130\%, \rho = -0.2$. The dashed curve of the same colour indicates the errors of the corresponding approximation.

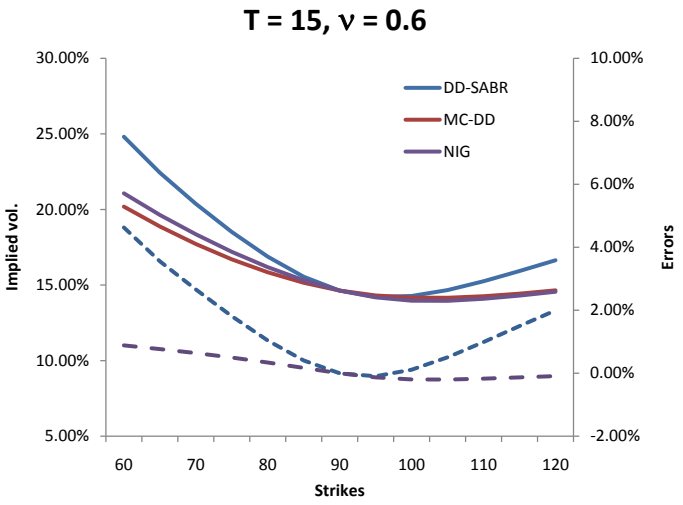

Figure 4.4: Effect of high $\nu^2 T$ (stress volatility regime) on the NIG approximation. Parameters: $\beta = 0.5, \nu = 0.6, S_0 = 90, \sigma_0 = 130\%, \rho = -0.2$. The dashed curve of the same colour indicates the errors of the corresponding approximation.

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 3 *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 4.2, we display the mass assigned to negative rates for both approximations and compare with*

exact (MC-DD) results for the cases considered in section 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 4.2: Probability mass assigned to the negative rates region for the four cases considered in figure 2.5.

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 paper are the SABR model and the DD-SABR model. In section 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 paper offers a potentially good approach to other SDE formulations which are more capable of exploring to the whole term structure of smiles.

References

Barndorff-Nielsen, O. E. (1997). Normal Inverse Gaussian Distributions and Stochastic Volatility Modelling. *Scandinavian Journal of Statistics*, *24(1)*, 1-13.

Berestycki, H., Busca, J., & Florent, I. (2004). Computing the implied volatility in stochastic volatility models. *Comm. Pure Appl. Math.*, *57(10)*, 1352-1373.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, *81(3)*, 637-659.

Cox, J. C. (1996). The constant elasticity of variance option pricing model. *Journal of Portfolio Management*, *23*, 15-17.

Hagan, P., Kumar, D., Lesniewski, A., & Woodward, D. (2002). Managing smile risk. *Wilmott Magazine*, *1(8)*, 84-108.

Hagan, P., Lesniewski, A., & Woodward, D. (2005). Probability distribution in the SABR model of stochastic volatility. *Preprint.*. (Available at http://lesniewski.us/papers/working/ProbDistrForSABR.pdf)

Henry-Labordere, P. (2005). A general asymptotic implied volatility for stochastic volatility models. *SSRN eLibrary.*. (Available at http://papers.ssrn.com/sol3/papers.cfm?abstract$_id = 698601$)

Johnson, S., & Nonas, B. (2009). *Arbitrage-free construction of the swaption cube.* http://ssrn.com/abstract=1330869.

Joshi, M., & Rebonato, R. (2003). A displaced diffusion stochastic volatility LIBOR market model: motivation, definition and implementation. *Quantitative Finance*, *3(6)*, 458-469.

Karatzas, I., & Shreve, S. (1991). *Brownian motion and stochastic calculus, 2nd edition.* Springer-Verlag, Berlin, Heidelberg and New York.

Larsson, K. (2010). *Dynamic Extensions and Probabilistic Expansions of the SABR model.* http://ssrn.com/paper=1536471.

Marris, D. (1999). *Financial option pricing and skewed volatility.* Unpublished master's thesis, University of Cambridge.

Mitra, S. (2010). *A One Factor Approximation for the SABR model.* Unpublished master's thesis, Warwick Business School, University of Warwick.

Obloj, J. (2008). Fine-tune your smile: Correction to hagan et al. *Wilmott Magazine*, *35*, 102-104.

Paulot, L. (2009). *Asymptotic Implied Volatility at the Second Order with Application to the SABR model.* http://ssrn.com/abstract=1413649.

Piterbarg, V. (2005). Stochastic volatility model with time-dependent skew. *Applied Mathematical Finance*, *12:2*, 147-185.

Rebonato, R. (2002). *Modern pricing of interest-rate derivatives.* Princeton University Press.

Rubinstein, M. (1983). Displaced diffusion option pricing. *Journal of Finance*, *38(1)*, 213-217.

Schroder, M. (1989). Computing the constant elasticity of variance option pricing formula. *Journal of Finance*, *44(1)*, 211-219.

Svoboda-Greenwood, S. (2009). Displaced diffusion as an approximation of the constant elasticity of variance. *Applied Mathematical Finance*, *16(3)*, 269-286.

Wu, Q. (2010). Series Expansion of the SABR Joint Density. *Mathematical Finance.*. (Available at http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9965.2010.00460.x/abstract)

A Distribution of F_T under the Log-Normal SABR model

The SDE of the Log-Normal SABR model is

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

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

$$\ln F_T = \ln F_0 + \frac{\rho}{\nu}(\sigma_T - \sigma_0) - \frac{1}{2} V_T + \sqrt{1-\rho^2} \int_0^T \sigma_t d\hat{W}_t, \tag{A.0.1}$$

where $V_T := \int_0^T \sigma_t^2 dt$. Let $M_t := \int_0^t \sigma_u d\hat{W}_u$ and $\mathcal{F}_T = \sigma(Z_u : 0 \le u \le T)$ 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

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

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

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

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 (3.1.2) and (3.1.4) respectively.

B Proof of Proposition 1: conditional moments of the realized variance V_T

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

$$Z_T = \frac{\ln(\sigma_T/\sigma_0) + \frac{1}{2}\nu^2 T}{\nu}. \tag{B.0.2}$$

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

$$Z_t|Z_T \triangleq Z_T\frac{t}{T} + (B_t - \frac{t}{T}B_T); \qquad 0 \le t \le T \tag{B.0.3}$$

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 & Shreve (1991)). It then follows from equation (B.0.3) that

$$Z_t|\sigma_T = Z_t|Z_T \sim \mathcal{N}\left(\frac{t}{T}Z_T, t - \frac{t^2}{T}\right), \tag{B.0.4}$$

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

$$\mathrm{Cov}(Z_t, Z_s|\sigma_T) = t \wedge s - \frac{ts}{T}. \tag{B.0.5}$$

B.1 First conditional moment of V_T

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

$$\begin{aligned}
\mathbb{E}(V_T \mid \sigma_T) &= \mathbb{E}\left[\int_0^T \exp\left(2\ln\sigma_0 - \nu^2 t + 2\nu Z_t\right) dt \mid \sigma_T\right] \\
&= \sigma_0^2 \exp\left(\frac{\left[\frac{\nu\sqrt{T}}{2} + \frac{Z_T}{\sqrt{T}}\right]^2}{2}\right)\sqrt{\frac{\pi T}{2}}\nu^{-1} \\
&\quad \times \left[\Phi\left(\frac{3\nu\sqrt{T}}{2} - \frac{Z_T}{\sqrt{T}}\right) - \Phi\left(-\frac{\nu\sqrt{T}}{2} - \frac{Z_T}{\sqrt{T}}\right)\right],
\end{aligned} \tag{B.1.1}$$

where Φ is the cumulative normal distribution function. Plugging back (B.0.2) to (B.1.1), we obtain

$$\mathbb{E}(V_T \mid \sigma_T) = \frac{\sigma_0^2\sqrt{T}}{2\nu}\frac{\left[\Phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}} + \nu\sqrt{T}\right) - \Phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}} - \nu\sqrt{T}\right)\right]}{\phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}} + \nu\sqrt{T}\right)}, \tag{B.1.2}$$

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

B.2 Second conditional moment of V_T

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

$$\begin{aligned}
\mathbb{E}(V_T^2 \mid \sigma_T) &= \mathbb{E}\left[2\int_0^T\int_0^t \sigma_t^2\sigma_s^2 ds dt \mid \sigma_T\right] \\
&= 2\sigma_0^4\int_0^T\int_0^t \exp\left(-\nu^2(t+s) + 2\nu\frac{t+s}{T}Z_T + 2\nu^2\left(t - \frac{t^2}{T} + s - \frac{s^2}{T} + 2s - 2\frac{ts}{T}\right)\right) ds dt \\
&= 2\sigma_0^4\int_0^T \exp(-4\nu^2 t)\int_0^t \exp\left(-\frac{4\nu^2\frac{(t+s)^2}{T} - 2\frac{2\nu(t+s)}{\sqrt{T}}\left(\frac{5\nu\sqrt{T}}{2} + \frac{Z_T}{\sqrt{T}}\right)}{2}\right) ds dt.
\end{aligned} \tag{B.2.1}$$

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 (B.2.1), we have that

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

The above integral can be evaluated by integration by parts

$$\begin{aligned}
\int_0^T \exp(-4\nu^2 t)\Phi(g_1(t))dt &= \left[-\frac{\exp(-4\nu^2 T)}{4\nu^2}\Phi\left(\frac{-Z_T}{\sqrt{T}} + \frac{3\nu}{2}\sqrt{T}\right) + \frac{1}{4\nu^2}\Phi\left(\frac{-Z_T}{\sqrt{T}} - \frac{5\nu}{2}\sqrt{T}\right)\right] \\
&\quad + \frac{1}{\nu\sqrt{T}}\exp\left(\frac{\left(\frac{3\nu\sqrt{T}}{2} + \frac{Z_T}{\sqrt{T}}\right)^2}{2} - \frac{\left(\frac{5\nu\sqrt{T}}{2} + \frac{Z_T}{\sqrt{T}}\right)^2}{2}\right) \\
&\quad \times \int_0^T \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\left(\frac{4\nu t - Z_T}{\sqrt{T}} - \frac{3\nu}{2}\sqrt{T}\right)^2}{2}\right) dt. \qquad \text{(B.2.2)}
\end{aligned}$$

By change of variable $v = \frac{4\nu t - Z_T}{\sqrt{T}} - \frac{3\nu}{2}\sqrt{T}$ in (B.2.2), we find that

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

Similarly,

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

Putting all the pieces together we obtain

$$\begin{aligned}\mathbb{E}(V_T^2 \mid \sigma_T) &= -\frac{\sigma_0^4\sqrt{T}}{4\nu^3}\left(1+e^{2\ln(\sigma_T/\sigma_0)}\right)\frac{\left[\Phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}}+\nu\sqrt{T}\right)-\Phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}}-\nu\sqrt{T}\right)\right]}{\phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}}+\nu\sqrt{T}\right)} \\ &\quad +\frac{\sigma_0^4\sqrt{T}}{4\nu^3}\frac{\left[\Phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}}+2\nu\sqrt{T}\right)-\Phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}}-2\nu\sqrt{T}\right)\right]}{\phi\left(\frac{\ln(\sigma_T/\sigma_0)}{\nu\sqrt{T}}+2\nu\sqrt{T}\right)}. \qquad \text{(B.2.3)}\end{aligned}$$

C 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

$$\begin{aligned}\mu(\sigma_T) &= \mathbb{E}(s_T|\sigma_T) \\ &= \ln F_0 + \frac{\rho}{\nu}(\sigma_T-\sigma_0) - \frac{1}{2}\mathbb{E}(V_T|\sigma_T) + \sqrt{1-\rho^2}\mathbb{E}(V_T^{\frac{1}{2}}G|\sigma_T). \qquad \text{(C.0.4)}\end{aligned}$$

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

$$\mu(\sigma_T) = \ln F_0 + \frac{\rho}{\nu}(\sigma_T-\sigma_0) - \frac{1}{2}\mathbb{E}(V_T|\sigma_T). \qquad \text{(C.0.5)}$$

The **conditional variance** of s_T:

$$\begin{aligned}\eta^2(\sigma_T) &= \mathrm{Var}(s_T|\sigma_T) \\ &= \frac{1}{4}\mathrm{Var}(V_T|\sigma_T) - \sqrt{1-\rho^2}\mathrm{Cov}(V_T, V_T^{\frac{1}{2}}G|\sigma_T) + (1-\rho^2)\mathrm{Var}(V_T^{\frac{1}{2}}G|\sigma_T). \qquad \text{(C.0.6)}\end{aligned}$$

Similarly, the covariance term in (C.0.6) 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 (C.0.6) 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

$$\eta^2(\sigma_T) = \frac{1}{4}\mathbb{E}(V_T^2|\sigma_T) - \frac{1}{4}[\mathbb{E}(V_T|\sigma_T)]^2 + (1-\rho^2)\mathbb{E}(V_T|\sigma_T). \tag{C.0.7}$$

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

D 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:

$$\begin{aligned} dF_t &= \hat{\sigma}_t C(F_t) dW_t, \\ d\hat{\sigma}_t &= \nu \hat{\sigma}_t dZ_t, \\ dZ_t dW_t &= \rho dt, \end{aligned}$$

where the function $C(u)$ is is assumed to be positive, smooth and integrable around 0:

$$\int_0^x \frac{du}{C(u)} < \infty, \qquad x > 0.$$

Equation (B.65) in appendix B of Hagan et al. (2002) yields the equivalent Black implied volatility for the above model:

$$\begin{aligned} \sigma_B(K, F_0) &= \frac{\hat{\sigma}_0 \ln F_0/K}{\int_K^{F_0} \frac{du}{C(u)}} \left(\frac{z}{x(z)}\right) \times \\ &\left\{1 + \left[\frac{2\gamma_2 - \gamma_1^2 + \frac{1}{F_{av}^2}}{24}\hat{\sigma}_0^2 C^2(F_{av}) + \frac{1}{4}\rho\nu\hat{\sigma}_0\gamma_1 C(F_{av}) + \frac{2-3\rho^2}{24}\nu^2\right] T + \ldots\right\}. \end{aligned} \tag{D.0.8}$$

Here

$$F_{av} = \sqrt{F_0 K}, \tag{D.0.9}$$

$$\gamma_1 = \frac{C'(F_{av})}{C(F_{av})}, \tag{D.0.10}$$

$$\gamma_2 = \frac{C''(F_{av})}{C(F_{av})}, \tag{D.0.11}$$

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).

$$z = \frac{\nu}{\hat{\sigma}_0}\int_K^{F_0}\frac{du}{C(u)}. \tag{D.0.12}$$

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 paper. 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 (D.0.10), (D.0.11) and (D.0.12) 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 (D.0.8), we get:

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

We can simplify this formula by expanding[§]

$$\begin{aligned} (F_0+\theta)-(K+\theta) &= \sqrt{(F_0+\theta)(K+\theta)}\ln\frac{F_0+\theta}{K+\theta}\left\{1+\frac{1}{24}\ln^2\frac{F_0+\theta}{K+\theta}+\frac{1}{1920}\ln^4\frac{F_0+\theta}{K+\theta}+\dots\right\}, \\ F_0-K &= \sqrt{F_0K}\ln\frac{F_0}{K}\left\{1+\frac{1}{24}\ln^2\frac{F_0}{K}+\frac{1}{1920}\ln^4\frac{F_0}{K}+\dots\right\}. \end{aligned}$$

Hence, the implied volatility formula now reads

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

[§]We use Hagan's technique with the assumption that the strike K is not so far away from F_0

For the special case of ATM options, we first take the limit

$$\lim_{K\to F_0}\frac{\ln F_0/K}{\ln(F_0+\theta)/(K+\theta)}=\frac{F_0+\theta}{F_0}, \tag{D.0.13}$$

and hence the formula reduces to

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