UNIVERSAL SMILES *

Article *in* Wilmott · July 2016
DOI: 10.1002/wilm.10523

Patrick S Hagan
Gorilla Science

Andrew Lesniewski
City University of New York - Bernard M. Baruch College

UNIVERSAL SMILES*

PATRICK S. HAGAN†, DEEP KUMAR‡, ANDREW S. LESNIEWSKI§, AND DIANA E. WOODWARD¶

Abstract. The SABR model has two variables, the forward asset price $\tilde{F}(T)$, and the local volatiliity $\tilde{A}(T)$. A singular perturbaton analysis has shown that the marginal density $Q(T,F)$, defined by

$$Q(T,F)dF = \text{prob}\left\{F < \tilde{F}(T) < F + dF\right\},$$

can be found through $O(\varepsilon^2)$ by solving a one dimensional *effective forward equation* of the form

$$Q_T = \tfrac{1}{2}\varepsilon^2 a^2 \left\{\left[1 + 2\varepsilon bz/\alpha + \varepsilon^2 cz/\alpha^2\right] C^2(F)Q\right\}_{FF},$$

where $b = \rho\nu$, $c = \nu^2$, and

$$z = \frac{1}{\varepsilon}\int_f^F \frac{dF'}{C(F')}.$$

This reduces the valuation of European options to one spatial dimension from two.

Recently, similar asymptotic analyses have shown that *for all commonly used stochastic volatility models*, the marginal density $Q(T_{ex},F)$ can be obtained through $O\left(\varepsilon^2\right)$ *by solving the same 1-d effective forward equation.* The only differences are in the formulas for the coefficients b and c in terms of each model's fundamental parameters. These stochastic volatility models include the Heston and generalized Heston models, the mean reverting SABR (λ-SABR) and dynamic SABR models, the exponential volatility models, ZABR-like models, cross FX SABR models, and the SABR models for baskets and spreads.

Here we analyze the above "universal" effective forward equation, and obtain explicit asymptotic formulas for the implied volatilities of European option for all the above models. These new formulas reduce to the original SABR implied vol formulas under moderate conditions, but are much more accurate in extreme situtations.

1. SABR.

1.1. The effective forward equation. Consider the standard SABR model[1] for an asset's forward price $\tilde{F}(T)$,

$$d\tilde{F} = \varepsilon\tilde{A}C(\tilde{F})d\tilde{W}_1, \tag{1.1a}$$

$$d\tilde{A} = \varepsilon\nu\tilde{A}d\tilde{W}_2, \tag{1.1b}$$

$$d\tilde{W}_1 d\tilde{W}_2 = \rho dT. \tag{1.1c}$$

The backbone function $C(F)$ is most commonly taken to be an offset CEV factor,

$$C(\tilde{F}) = \left(\tilde{F} + o\right)^\beta. \tag{1.2}$$

For many markets, the offset o is zero. In rates, however, the offset is often taken to be 50-150bps to allow the foward rate $\tilde{F}$ to be slightly negative. Here we treat $\varepsilon \ll 1$ as a small parameter, finding asymptotic

*The information contained herein reflects, as of the date hereof, the views and opinions of the authors and sources that they believe to be reliable and does not necessarily reflect the views of AVM, L.P. or its affiliates, employees or principals (collectively, "AVM"). The authors' views expressed herein may change at any time subsequent to the date of issue hereof and neither the authors nor AVM, undertake to advise you of any changes in the views expressed herein. No representation or warranty (express or implied) is given as to the accuracy or completeness of the information contained herein. These materials are provided for informational purposes only, and under no circumstances may any information contained herein by construed as investment advice or an offer to sell or a solicitation of an offer to purchase (or any marketing in connection thereof) any security.

†AVM LP; 777 Yamato Road; Boca Raton FL 33431
‡AVM LP; 777 Yamato Road; Boca Raton FL 33431
§Dept of Mathematics; Baruch College; One Baruch Way; New York, NY 10010
¶Gorilla Science, 7700 NE Palm Way, Boca Raton, FL 33487

formulas for the implied volatilitilities of European options. After obtaining the final formulas, we can then set ε to 1.[1]

Let the initial conditions be $\tilde{F}(0) = f$ and $\tilde{A}(0) = \alpha$, and define the probability density of being at F, A at time T to be

$$p(T, F, A; f, \alpha) = \mathrm{E}\left\{ \delta(\tilde{F}(T) - F)\delta(\tilde{A}(T) - A) \;\Big|\; \tilde{F}(0) = f,\ \tilde{A}(0) = \alpha \right\}. \tag{1.3}$$

Also define the marginal (reduced) density by

$$Q(T, F; f, \alpha) \equiv \mathrm{E}\left\{ \delta(\tilde{F}(T) - F) \;\Big|\; \tilde{F}(0) = f, \tilde{A}(0) = \alpha) \right\} \equiv \int_0^\infty p(T, F, A; f, \alpha) dA. \tag{1.4}$$

In [4], [5] it was shown that the marginal density can be found through $O(\varepsilon^2)$ by solving the one dimensional *effective forward equation*[2]

$$Q_T = \tfrac{1}{2}\varepsilon^2\alpha^2 e^{\varepsilon^2\rho\nu\alpha\Gamma_0 T}\left[\left(1 + 2\varepsilon\rho\nu z/\alpha + \varepsilon^2\nu^2 z^2/\alpha^2\right) C^2(F) Q\right]_{FF}, \tag{1.5a}$$

$$Q \to \delta\left(F - f\right) \qquad \text{as } T \to 0, \tag{1.5b}$$

where

$$z = \frac{1}{\varepsilon}\int_f^F \frac{dF'}{C(F')}, \tag{1.5c}$$

and $\Gamma_0 = C'(f)$.

In [4] we solved the effective forward equation numerically, and then integrated,

$$V_{call}(T_{ex}, K) = \int \left(F - K\right)^+ Q(T_{ex}, F)\, dF, \tag{1.6a}$$

$$V_{put}(T_{ex}, K) = \int \left(K - F\right)^+ Q(T_{ex}, F)\, dF, \tag{1.6b}$$

to find the values of European options with expiry T_{ex} and strike K. This approach is expedient, since the effective forward equation has only one spatial dimension.

Here we take a different approach, an analytical approach, to obtaining European option prices. This approach requires the effective forward equation to be stationary, so we first need to eliminate the time dependent factor $e^{\varepsilon^2\rho\nu\alpha\Gamma_0 T}$. To obtain the European option prices for any given expiry date T_{ex}, we define a new time variable,

$$T_{new}(T) = \frac{e^{\varepsilon^2\rho\nu\alpha\Gamma_0 T} - 1}{e^{\varepsilon^2\rho\nu\alpha\Gamma_0 T_{ex}} - 1} T_{ex} = T\left\{1 + \tfrac{1}{2}\varepsilon^2\rho\nu\alpha\Gamma_0\left(T - T_{ex}\right) + \cdots\right\}. \tag{1.7}$$

In the new time variable,

$$\frac{\varepsilon^2\rho\nu\alpha\Gamma_0 T_{ex}}{e^{\varepsilon^2\rho\nu\alpha\Gamma_0 T_{ex}} - 1} Q_{T_{new}} = \tfrac{1}{2}\varepsilon^2\alpha^2\left[\left(1 + 2\varepsilon\rho\nu z/\alpha + \varepsilon^2\nu^2 z^2/\alpha^2\right) C^2(F) Q\right]_{FF}, \tag{1.8a}$$

$$Q \to \delta\left(F - f\right) \qquad \text{as } T_{new} \to 0. \tag{1.8b}$$

[1] Although this appears inconsistent, it is equivalent to non-dimensionalizing the model, effecting a small volatility expansion, and then reverting back to the original, dimensioned variables.

[2] Since the effective forward equation depends on f as well as F, it is not a true forward equation. That is, there is no local volatility model that has this equation as its forward Kolmogorov equation.

Since $T_{new}(T_{ex}) = T_{ex}$, the solution of 1.8a, 1.8b at $T_{new} = T_{ex}$ gives the marginal density $Q(T_{ex}, F)$ at our original expiry date. We now set

$$\alpha_{new} = \alpha \left(\frac{e^{\varepsilon^2 \rho\nu\alpha\Gamma_0 T_{ex}} - 1}{\varepsilon^2 \rho\nu\alpha\Gamma_0 T_{ex}} \right)^{1/2} = \alpha \left\{ 1 + \tfrac{1}{4}\varepsilon^2 \rho\nu\alpha\Gamma_0 T_{ex} + \cdots \right\}. \tag{1.9}$$

In terms of the new α and new T, the marginal density $Q(T_{ex}, F)$ is given by the solution of the time-independent equation

$$Q_T = \tfrac{1}{2}\varepsilon^2\alpha^2 \left[\left(1 + 2\varepsilon\rho\nu z/\alpha + \varepsilon^2\nu^2 z^2/\alpha^2\right) C^2(F) Q \right]_{FF} \qquad \text{for } 0 < T < T_{ex}, \tag{1.10a}$$

$$Q \to \delta(F - f) \qquad \text{as } T \to 0, \tag{1.10b}$$

through $O(\varepsilon^2)$, where

$$z = \frac{1}{\varepsilon} \int_f^F \frac{dF'}{C(F')}, \tag{1.10c}$$

and where we are omitting the *new* subscripts for simplicity. Note that by using α_{new} in

$$1 + 2\varepsilon \frac{\rho\nu}{\alpha} z + \varepsilon^2 \frac{\nu^2}{\alpha^2} z^2, \tag{1.11}$$

instead of using the orginal α, we are only introducing $O(\varepsilon^3)$ and smaller errors. These are negligible as we are only working through $O(\varepsilon^2)$.

If the volatility were not stochastic, so $\tilde{A}(T)$ was constant with value $\tilde{A}(0) = \alpha$, then we'd have a local volatility model

$$d\tilde{F} = \varepsilon\alpha C(\tilde{F}) d\tilde{W}, \tag{1.12a}$$

whose forward Kolmogorov equation (Fökker-Planck equation) would be

$$Q_T = \tfrac{1}{2}\varepsilon^2\alpha^2 \left[C^2(F) Q \right]_{FF}. \tag{1.12b}$$

Comparing this to eq. 1.10a shows that the effect of stochastic volatility is to turn the factor α^2 into the factor $\alpha^2\left(1 + 2\varepsilon\rho\nu z/\alpha + \varepsilon^2\nu^2 z^2/\alpha^2\right)$. This factor arises because the value of an option is nonlinear in the local volatility $\tilde{A}(T)$ as well as the forward $\tilde{F}(T)$. Since the local volatility is itself stochastic, the option would systematically gain or lose money, depending on the sign of the option's gamma with respect to the volatility (vol gamma). The factor $\alpha^2\left(1 + 2\varepsilon\rho\nu z/\alpha + \varepsilon^2\nu^2 z^2/\alpha^2\right)$ increases or decreases theta (the daily carry $\partial V/\partial T$), to balance the systematic gain or loss caused by fluctuations in the volatility, thus rendering the pricing arbitrage free.

1.2. Implied normal vols. In appendix A we use singular perturbation techniques to analyze equations 1.10a-1.10c. This analysis yields the marginal density $Q(T_{ex}, F)$ and European option prices 1.6a, 1.6b through $O(\varepsilon^2)$. These prices then enable us to derive explicit formulas for the implied normal vols[3] (*aka*,

[3] Implied Black (log normal) vols can be obtained from σ_N via the conversion formula:

$$\sigma_B(T_{ex}, K) = \sigma_N(T_{ex}, K) \frac{\log f/K}{f - K} \left\{ 1 + \tfrac{1}{24} \frac{\sigma_N^2 T_{ex}}{fK} + \cdots \right\}$$

the absolute or Bachelier vols), which are again accurate through $O\left(\varepsilon^2\right)$. Specifically, the implied normal vol for a European option with expiry T_{ex} and strike K is

$$\sigma_N\left(T_{ex},K\right)=\varepsilon\alpha\cdot\frac{K-f}{\displaystyle\int_f^K\frac{dF'}{C\left(F'\right)}}\cdot\frac{\zeta}{Y\left(\zeta\right)}\cdot\left\{\begin{array}{ll}1+\varepsilon^2\theta\left(\zeta\right)T_{ex}+\cdots & \text{if } \theta\geq 0\\ \dfrac{1}{1-\varepsilon^2\theta\left(\zeta\right)T_{ex}+\cdots} & \text{if } \theta<0\end{array}\right., \tag{1.13a}$$

where the stochastic volatility terms are

$$\zeta=\frac{\nu}{\alpha}\int_f^K\frac{dF'}{C\left(F'\right)},\qquad Y\left(\zeta\right)=\log\frac{\rho+\zeta+\sqrt{1+2\rho\zeta+\zeta^2}}{1+\rho}, \tag{1.13b}$$

and where the higher order term is

$$\theta\left(\zeta\right)=\frac{\nu^2}{24}\left\{-1+3\frac{\rho+\zeta-\rho E\left(\zeta\right)}{Y\left(\zeta\right)E\left(\zeta\right)}\right\}+\frac{\Delta_0\alpha^2}{6}\left\{1-\rho^2+\frac{\left(\rho+\zeta\right)E\left(\zeta\right)-\rho}{Y\left(\zeta\right)}\right\}, \tag{1.13c}$$

with

$$E\left(\zeta\right)=\sqrt{1+2\rho\zeta+\zeta^2},\qquad \Delta_0=\tfrac{1}{4}C(f)C''(f)-\tfrac{1}{8}\left[C'\left(f\right)\right]^2. \tag{1.13d}$$

The most common case in practice is an offset CEV backbone, $C\left(F\right)=\left(F+o\right)^{\beta}$. For this special case,

$$\frac{K-f}{\displaystyle\int_f^K\frac{dF'}{C\left(F'\right)}}=\frac{\left(1-\beta\right)\left(K-f\right)}{\left(K+o\right)^{1-\beta}-\left(f+o\right)^{1-\beta}}, \tag{1.14a}$$

and

$$\zeta=\frac{\nu}{\alpha}\frac{\left(K+o\right)^{1-\beta}-\left(f+o\right)^{1-\beta}}{1-\beta},\qquad \Delta_0=-\tfrac{1}{8}\frac{\beta\left(2-\beta\right)}{\left(f+o\right)^{2-2\beta}}. \tag{1.14b}$$

In the implied normal vol formula 1.13a, the main factors can be written as

$$\varepsilon\alpha\cdot\frac{1}{\displaystyle\frac{1}{K-f}\int_f^K\frac{dF'}{C\left(F'\right)}}\cdot\frac{1}{\displaystyle\frac{1}{\zeta}\int_0^{\zeta}\frac{d\zeta_1}{\sqrt{1+2\rho\zeta_1+\zeta_1^2}}}, \tag{1.15}$$

since

$$Y\left(\zeta\right)=\int_0^{\zeta}\frac{d\zeta_1}{\sqrt{1+2\rho\zeta_1+\zeta_1^2}}. \tag{1.16}$$

The factor $\varepsilon\alpha$ determines the overall volatility level. The next factor is the main effect of the backbone function, $C(F)$. For strikes K not too far from f, this factor is approximately the average $\frac{1}{2}\left\{C\left(f\right)+C\left(K\right)\right\}$. The third factor $\zeta/Y\left(\zeta\right)$ is the main stochastic volaility effect. It changes the option's value to compensate for the gamma with respect to the volatility. It too represents an average, albeit an unusual one, of the stochastic volatility effect, $1+2\varepsilon\rho\nu z/\alpha+\varepsilon^2\nu^2z^2/\alpha^2$. Finally, the last factor

$$\left\{\begin{array}{ll}1+\varepsilon^2\theta\left(\zeta\right)T_{ex}+\cdots & \text{if } \theta\geq 0\\ \dfrac{1}{1-\varepsilon^2\theta\left(\zeta\right)T_{ex}+\cdots} & \text{if } \theta<0\end{array}\right. \tag{1.17}$$

contains all the higher order effects.

We could have used either $\left\{1+\varepsilon^2\theta T_{ex}+\cdots\right\}$ or $\left\{1-\varepsilon^2\theta T_{ex}+\cdots\right\}^{-1}$ exclusively for all θ, since they are asymptotically equivalent through $O(\varepsilon^2)$. However, in extreme cases, where $\varepsilon^2\theta T_{ex}$ is not small, making the choice in eq. 1.17 gives a more robust formula for the implied normal vol. As a side benefit, it also appears to be more accurate than using either form exclusively.

The implied volatility formula in eqs. 1.13a - 1.13d is very similar to the implied volatility formula derived in the original SABR paper[1]. In fact, making the replacement

$$\theta(\zeta) \to \theta(0) = \tfrac{2-3\rho^2}{24}\nu^2 + \tfrac{1}{3}\Delta_0\alpha^2 \tag{1.18}$$

essentially reduces the new implied vol formula to the original SABR formula.

Caution. In using the implied volatility formula 1.13a - 1.13d, note that the α in these equations is the new α,

$$\alpha_{new} = \alpha_0\left\{1+\tfrac{1}{4}\varepsilon^2\rho\nu\alpha_0\Gamma_0 T_{ex}+\cdots\right\}, \tag{1.19}$$

where $\alpha_0 = \tilde{A}(0)$ is the original, model α. See eq. 1.9. The original SABR formulas in [1],

$$\sigma_N(T_{ex},K) = \varepsilon\alpha\cdot\frac{K-f}{\int_f^K \frac{dF'}{C(F')}}\cdot\frac{\zeta}{Y(\zeta)}\cdot\left\{1+\varepsilon^2\theta(0)T_{ex}+\cdots\right., \tag{1.20}$$

use the original α_0 for α. Consequently, the original formulas have

$$\theta(0) = \tfrac{1}{4}\rho\nu\alpha\Gamma_0 + \tfrac{2-3\rho^2}{24}\nu^2 + \tfrac{1}{3}\Delta_0\alpha^2. \tag{1.21}$$

In our new formulas, there is no term $\frac{1}{4}\rho\nu\alpha\Gamma_0$, as it has already been included in the leading factor

$$\varepsilon\alpha_{new}\cdot\ \cdots \equiv \varepsilon\alpha\left\{1+\tfrac{1}{4}\varepsilon^2\rho\nu\alpha\Gamma_0 T_{ex}+\cdots\right\}\cdot\ \cdots. \tag{1.22}$$

1.3. Examples. Figure 1.1 graphs the implied normal volatilities as a function of the strike K. The "exact" values have been obtained by numerically solving the effective forward equation 1.5a-1.5c by a Crank-Nicholson scheme, and then evaluating option prices 1.6a, 1.6b via numerical integration. The "original" implied volatilities are the implied volatitilities obtained from the classical SABR formulas, essentially 1.13a-1.13d with $\theta(\zeta)$ replaced by $\theta(0)$. See 1.18 above. The "new" values are the values obtained from formulas 1.13a-1.13d. Even though this is an admittedly extreme case (the "small parameter" $\varepsilon^2\nu^2T_{ex}$ is nearly 3!), the new SABR formulas remain reasonably accurate for all strikes except very low strikes. In more moderate situations, there is much less difference between the new, original, and exact implied vols.

Figure 1.2 graphs the values of the out-of-the-money calls and puts, again comparing the numerically obtained exact values with the values obtained from the classical SABR formula and the new SABR formula. Again, the exact, the original SABR, and the new SABR values are usually much closer together.

The terminal marginal probability density can be implied from the European option prices, as

$$\frac{\partial^2}{\partial^2 K}V_{call}(T_{ex},K) = \frac{\partial^2}{\partial^2 K}\int \left(F-K\right)^+ Q(T_{ex},F)\,dF = Q(T_{ex},K). \tag{1.23}$$

Figure 1.3 compares the densities derived from the exact values with the densities derived from the original and new SABR formulas. The exact probability density obtained by numerically solving the effective forward equation is always positive, courtesy of the maximum principle[12]. This ensures that the numerical solution is arbitrage free[4], [13], [14]. The original and new SABR formulas lead to negative probability densities, and are thus not arbitrage free, for very low strikes for the case shown.

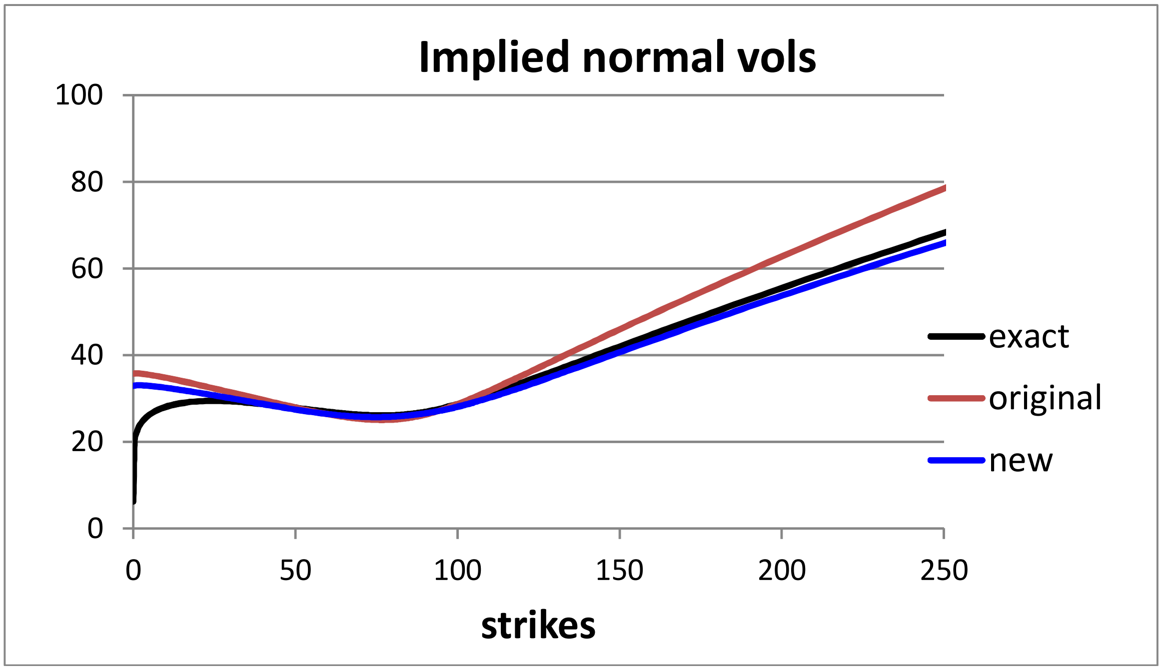

FIG. 1.1. *Implied volatilties for $f = 100$, $T_{ex} = 3y$ and SABR parameters $\alpha = 7.91$, $\beta = 0.25$, $\rho = 50\%$, $\nu = 0.90$. Shown is the "exact" curve obtained by numerically solving the effective forward equation, with the standard SABR implied vol formula ("original") and the new formula.*

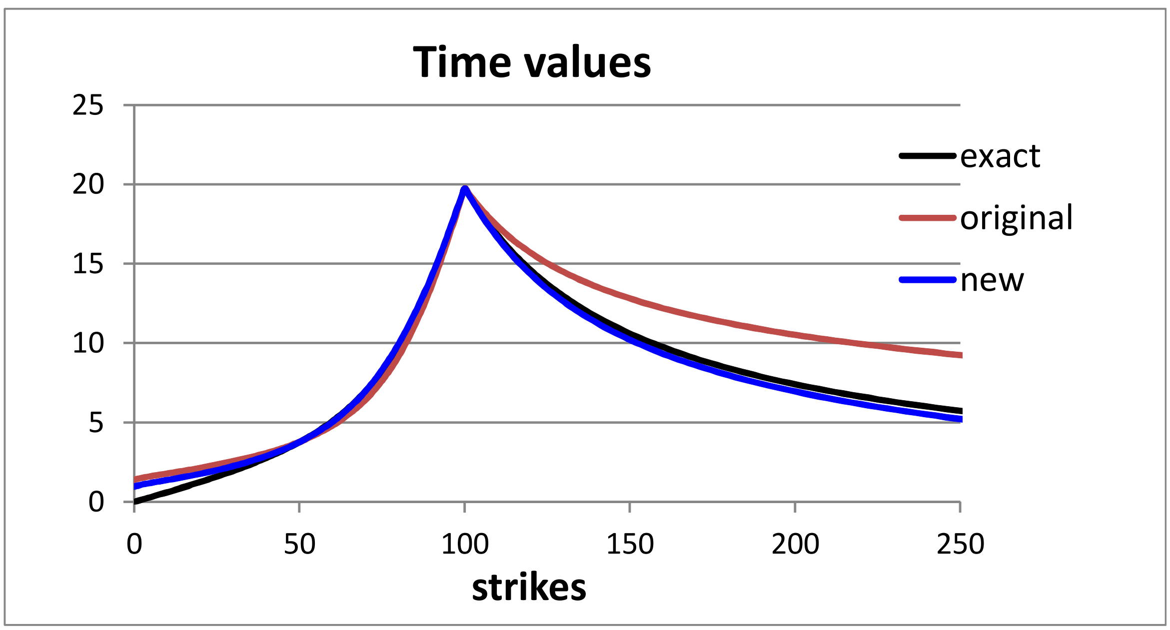

FIG. 1.2. *The values of out-of-the-money calls and puts under the SABR model.for $f = 100$ and $T_{ex} = 3y$. The SABR parameters are the same as in Figure 1.1. Shown are the numerically calculated exact values, along with the values obtained from the original and new SABR formulas.*

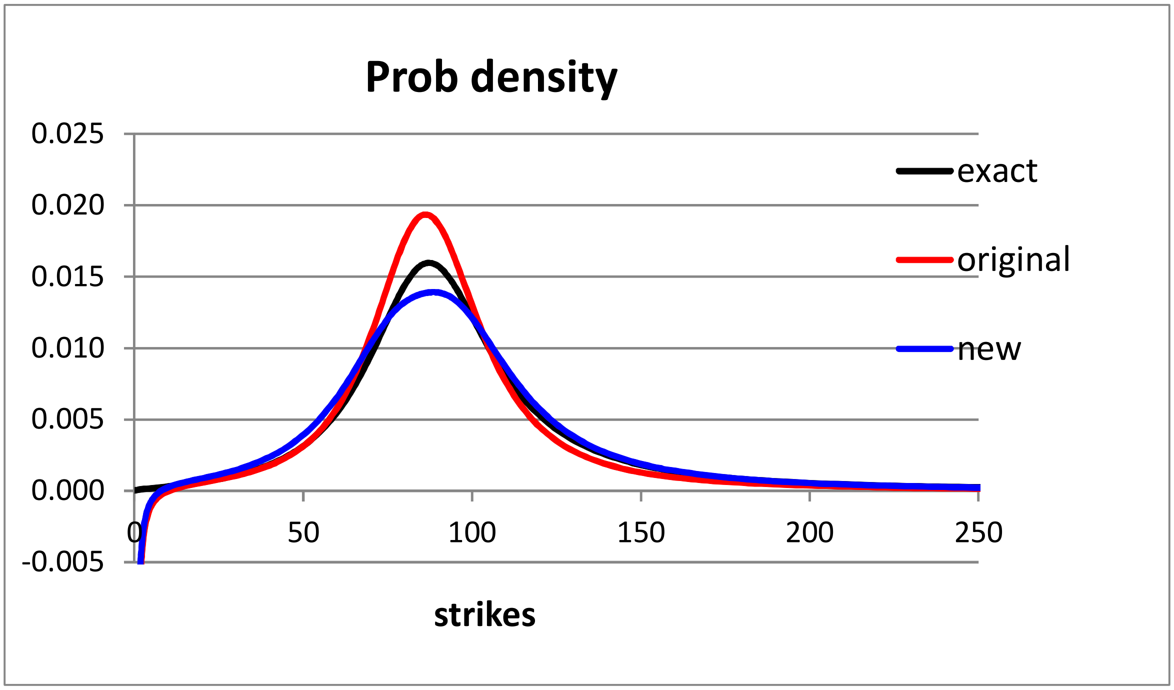

FIG. 1.3. *The probability density implied by the option prices for $f = 100$, $T_{ex} = 3\,$y. The SABR parameters are the same as in Figures 1.1 and 1.2. The density computed numerically (exact) remains positive for all K, but both implied vol formulas lead to negative densities for very low strikes K.*

1.4. Boundary corrections. In most stochastic volatilty models, the backbone function $C\left(F\right)$ has a zero at some point $F_{\min}$. This point acts as a barrier to diffusion. For example, the barrier is at $F_{\min} = -o$ for an offset CEV backbone, $C\left(F\right) = \left(F + o\right)^{\beta}$. It is nearly always assumed that the forward cannot breach this barrier, so $\tilde{F}(T) \geq F_{\min}$. All such boundaries were ignored in deriving the implied volatility formulas 1.13a-1.13d in Appendix A. Consequently, we can obtain better prices by accounting for the boundary when either the forward f or the strike K is near the boundary $F_{\min}$.

In [4] it is shown that these boundaries must have absorbing boundary conditions,

$$C^2(F)Q(T,F) \to 0 \text{ as } F \to F_{\min}, \tag{1.24}$$

in order for F to be a Martingale. We can incorporate this into our pricing by using the reflection principle[33]. This yields

$$V_{call}(T_{ex},K) = V^N_{call}(T_{ex},f,K;\sigma_N) - V^N_{put}(T_{ex},f,-K+2F_{\min};\sigma_N), \tag{1.25a}$$

$$V_{put}(T_{ex},K) = V^N_{put}(T_{ex},f,K;\sigma_N) - V^N_{put}(T_{ex},f,-K+2F_{\min};\sigma_N). \tag{1.25b}$$

Here

$$V^N_{call}(T_{ex},f,K;\sigma_N) = (f-K)\mathcal{N}\left(\frac{f-K}{\sigma_N T_{ex}^{1/2}}\right) + \sigma_N T_{ex}^{1/2}\mathcal{G}\left(\frac{f-K}{\sigma_N T_{ex}^{1/2}}\right), \tag{1.25c}$$

$$V^N_{put}(T_{ex},f,K;\sigma_N) = (K-f)\mathcal{N}\left(\frac{K-f}{\sigma_N T_{ex}^{1/2}}\right) + \sigma_N T_{ex}^{1/2}\mathcal{G}\left(\frac{K-f}{\sigma_N T_{ex}^{1/2}}\right), \tag{1.25d}$$

are the option prices under the normal model, and $\mathcal{N}(\cdot)$ is the cumulative normal distribution, $\mathcal{G}(\cdot)$ is the Gaussian density, and $\sigma_N(T_{ex}, K)$ is the implied volatility in 1.13a-1.13d. Thus, the first terms in 1.25a, 1.25b are the standard option prices for an implied normal vol of $\sigma_N(T_{ex}, K)$. The effect of the boundary is the second term, which subtracts the value of an out-of-the-money put from the option price. The value of this put exactly cancels the time value in the first term if either $f = F_{\min}$ or $K = F_{\min}$. However, as f and K increase away from the boundary, the put value rapidly decreases and quickly becomes negligible.

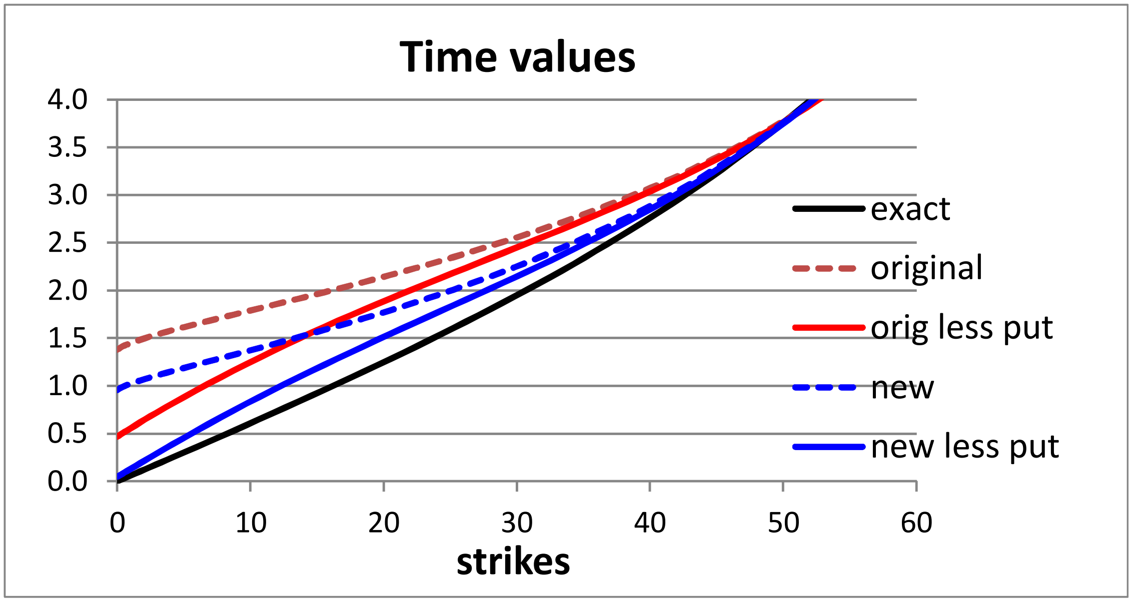

FIG. 1.4. *Time values for the same options as in Figs. 1.1-1.3, showing the effect of subtracting the put to account for the boundary condition. Only the low strike part of the smile is shown.*

Figure 1.4 presents option prices with and without the boundary put for low strikes. We see that including the boundary put substantially increases the accuracy for ultra low strikes, and that this correction decreases into insignificance as the strike K increases away from the boundary at $F_{\min} = 0$. This is reinforced in Figure 1.5, where the implied vols are shown for the same options, with and without the boundary put.

Even though including the boundary put yields very accurate option values, this does not ensure that these prices are arbitrage free for ultra low strikes. In fact, Figure 1.6 shows that even after subtracting the boundary put, the implied probability densities are negative for low enough strikes, at least for the case shown.

1.5. Curtailing the implied vols at extreme strikes. In the effective forward equation,

$$Q_T = \tfrac{1}{2}\varepsilon^2\alpha^2 \left[\left(1 + 2\varepsilon\rho\nu z/\alpha + \varepsilon^2\nu^2 z^2/\alpha^2\right) C^2(F)Q\right]_{FF}, \tag{1.26}$$

the factor

$$1 + 2\varepsilon\rho\nu z/\alpha + \varepsilon^2\nu^2 z^2/\alpha^2 \tag{1.27}$$

encapsulates the stochastic volatility effect (SVE) of the SABR model. This factor should not be too far from one, since z should be moderate and ε should be small. In practice, this factor may become unreasonably large, causing unreasonably large volatilities, for very high or very low strikes. To prevent this, trading desks

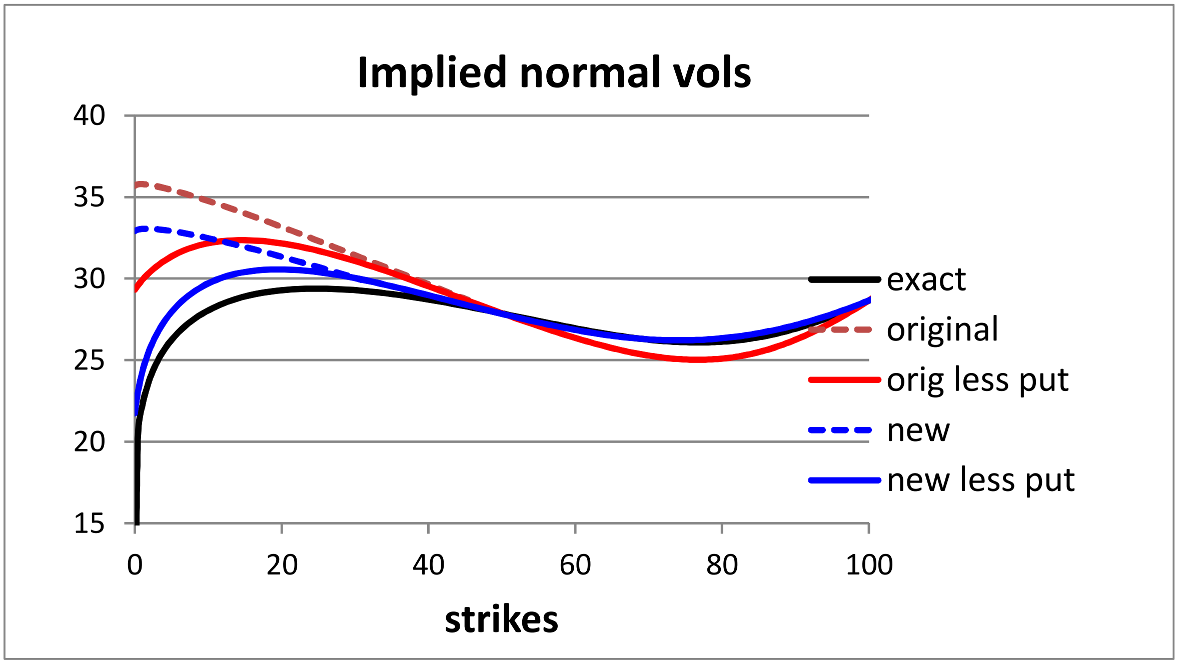

FIG. 1.5. *Implied vols for the same options as in Figs. 1.1-1.3, showing the effect of subtracting the put to account for the boundary condition. Only the low strike part of the smile is shown.*

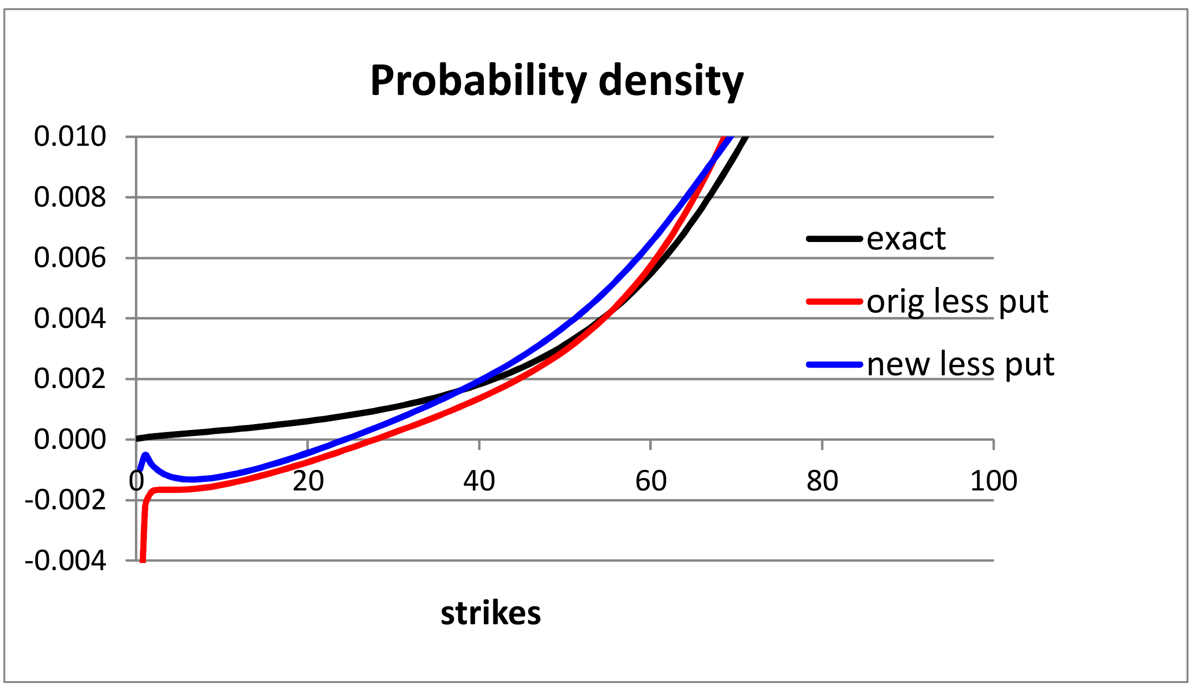

FIG. 1.6. *Implied probability densities for the same options as in Figs. 1.1-1.5. These densities are derived from the option values after subtracting the boundary put. Only the low strikes are shown.*

often adopt various *ad hoc* methods for curtailing implied vols at extreme strikes. For example, many use flat extrapolation of the implied normal volatilities beyond certain cut-off strikes. These pragmatic methods often lead to pricing artifacts, including arbitrage, near cut-off points.

We prefer to curtail the implied volatilities by directly capping the SVE. We do this by replacing the effective forward equation, eq. 1.26 by

$$Q_T = \tfrac{1}{2}\varepsilon^2\alpha^2\left[E_c^2(\varepsilon z)C^2(F)Q\right]_{FF}, \tag{1.28a}$$

where

$$E_c(\varepsilon z) = \min\left\{\sqrt{1+2\varepsilon\rho\nu z/\alpha+\varepsilon^2\nu^2z^2/\alpha^2}, E_{\max}\right\}. \tag{1.28b}$$

Here $E_{\max} > 1$ is the cap on the SVE. Eqs. 1.29a-1.30d below present the implied normal vols for the SABR model with a capped SVE. Even though these implied volatility formulas are complex, they are explicit and can be evaluated instantaneously. Moreover, the arguments in §3 show that capping the SVE does not introduce any arbitrages. This suggests that capping the SVE may be a potent method for managing CMS swaps, caps, and floors, instruments whose convexity corrections are notoriously sensitive to high strike vols[35]. Even relatively large caps, like $E_{\max} = 4$ or $E_{\max} = 5$, may be effective in moderating the high and low strike wings. See Figure 1.7.

The implied volatilities are derived in Appendix A for the effective forward equation 1.28a with an arbitrary, smooth, positive function $E_c(\varepsilon z)$. These formulas are then specialized to the capped $E_c(\varepsilon z)$ in 1.28b. This yields

$$\sigma(T_{ex}, K) = \varepsilon\alpha\cdot\frac{(K-f)}{\displaystyle\int_f^K\frac{dF}{C(F)}}\cdot\frac{\zeta}{Y(\zeta)}\cdot\begin{cases}1+\varepsilon^2\theta(\zeta)T_{ex} & \text{if } \theta\geq 0\\ \dfrac{1}{1-\varepsilon^2\theta(\zeta)T_{ex}} & \text{if } \theta<0\end{cases}, \tag{1.29a}$$

with

$$\zeta = \frac{\nu}{\alpha}\int_f^K\frac{dF}{C(F)}, \qquad E(\zeta) = \sqrt{1+2\rho\zeta+\zeta^2}, \tag{1.29b}$$

as before. Now, however,

$$Y(\zeta) = \begin{cases} Y_- - (\zeta_- - \zeta)/E_{\max} & \text{if } \zeta<\zeta_-,\\ \log\dfrac{\sqrt{1+2\rho\zeta+\zeta^2}+\rho+\zeta}{1+\rho} & \text{if } \zeta_-<\zeta<\zeta_+,\\ Y_+ + (\zeta-\zeta_+)/E_{\max} & \text{if } \zeta_+<\zeta.\end{cases} \tag{1.29c}$$

Here,

$$\zeta_\pm = -\rho\pm\sqrt{E_{\max}^2-1+\rho^2}, \qquad Y_\pm = \pm\log\frac{E_{\max}+\sqrt{E_{\max}^2-1+\rho^2}}{1\pm\rho} \tag{1.29d}$$

are the values of ζ, $Y(\zeta)$ at which $E(\zeta)$ hits $E_{\max}$. For $\zeta<\zeta_-$, the higher order correction term $\theta(\zeta)$ is

$$\begin{aligned}\theta(\zeta) = {}&\frac{\nu^2}{24Y(\zeta)}\left\{-Y_- - 3\frac{\sqrt{E_{\max}^2-1+\rho^2}}{E_{\max}} - 3\rho\right\}\\ &+\frac{\Delta_0\alpha^2}{6Y(\zeta)}\left\{2E_{\max}(\zeta-\zeta_-)+(1-\rho^2)Y_- - E_{\max}\sqrt{E_{\max}^2-1+\rho^2}-\rho\right\};\end{aligned} \tag{1.30a}$$

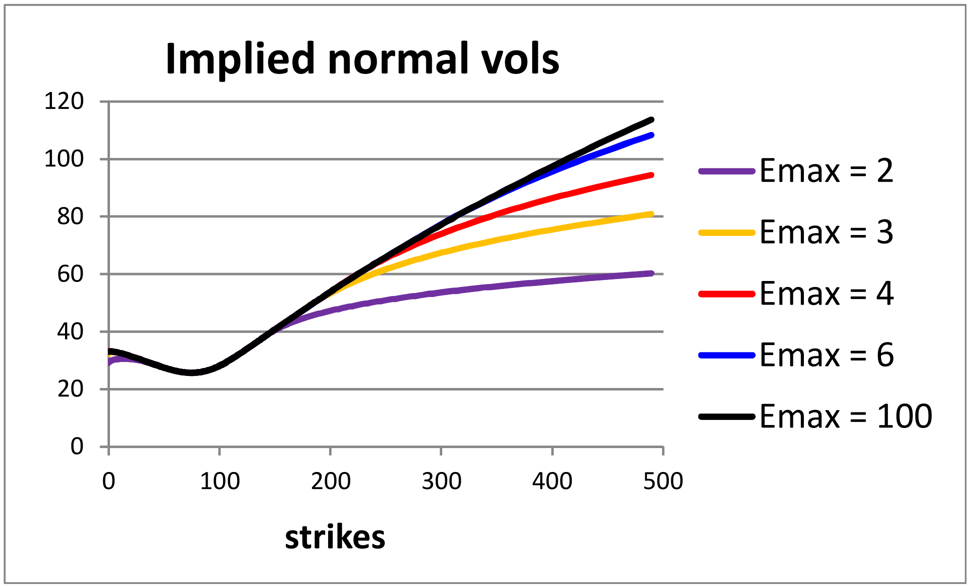

FIG. 1.7. *Shown are the implied volatilities for different SVE caps, $E_{\max}$. The option and SABR parameters are the same as Figs. 1.1-1.6.*

for $\zeta_- < \zeta < \zeta_+$ it is

$$\theta(\zeta) = \frac{\nu^2}{24}\left\{-1+3\frac{\rho+\zeta-\rho E(\zeta)}{Y(\zeta)\,E(\zeta)}\right\} + \frac{\Delta_0\alpha^2}{6}\left\{1-\rho^2+\frac{(\rho+\zeta)\,E(\zeta)-\rho}{Y(\zeta)}\right\}; \tag{1.30b}$$

and for $\zeta > \zeta_+$, it is

$$\begin{aligned}\theta(\zeta) = {}& \frac{\nu^2}{24Y(\zeta)}\left\{-Y_+ + 3\frac{\sqrt{E_{\max}^2-1+\rho^2}}{E_{\max}} - 3\rho\right\} \\ &+\frac{\Delta_0\alpha^2}{6Y(\zeta)}\left\{2E_{\max}(\zeta-\zeta_+) + (1-\rho^2)\,Y_+ + E_{\max}\sqrt{E_{\max}^2-1+\rho^2}-\rho\right\}.\end{aligned} \tag{1.30c}$$

Here

$$\Delta_0 = \tfrac{1}{4}C(f)C''(f) - \tfrac{1}{8}\left[C'(f)\right]^2. \tag{1.30d}$$

2. Universality of the smile formulas. Many popular stochastic volatility models have been analyzed recently using singular perturbation techniques [16]-[22]. Included are the dynamic and mean-reverting SABR (λ-SABR) models, the Heston and generalized Heston models, the exponential stochastic volatility model, the cross FX rate SABR model, and the SABR models for baskets and spreads. These models are set out in Appendix B. For each model, the marginal density

$$Q(T,F)dF = \text{prob}\left\{F < \tilde{F}(T) < F+dF\right\} \tag{2.1}$$

is defined, and the analysis shows that this marginal density satisfies an effective forward equation of the form

$$Q_T = \tfrac{1}{2}\varepsilon^2\alpha^2\sigma^2(T)\left[\left(1+2\varepsilon b(T)\,z/\alpha+\varepsilon^2 c(T)\,z^2/\alpha^2\right)C^2(F)Q\right]_{FF} \qquad \text{for } T>0, \tag{2.2a}$$

$$Q = \delta(F-f) \qquad \text{as } T\to 0, \tag{2.2b}$$

through $O\left(\varepsilon^2\right)$, where

$$z \equiv \frac{1}{\varepsilon}\int_f^F \frac{dF'}{C(F')}. \tag{2.2c}$$

The singular perturbation analyses determine the coefficients $\sigma(T)$, $b(T)$, and $c(T)$ in terms of the original model parameters, with the constant α usually chosen so that $\sigma(0) = 1$.

The effective forward equation has a *single* space dimension F, unlike the actual Kolmogorov forward equations of the original models, which have at least two space variables (the forward F and the volatility or variance A). Thus, the perturbation analyses reduce all the above multi-dimensional stochastic models to *the same one dimensional model*. Since the effective forward equation depends on both the backwards variable f and the forwards variable F, *there is no local volatility model* for which the effective forward equation is a true forward Kolmogorov (Fökker-Planck) equation.

We approach equations 2.2a-2.2c analytically in §2.1 below, using effective media results from [15]. We also briefly consider a numerical approach to solving these equations in §2.2. Before continuing, we caution that all derivations of the effective 1-d forward equations depend on a free space analysis of the forward Kolmogorov equations, and do not account for any imposed boundary conditions[11]. This is not a concern for pricing European options, as there are no imposed boundary conditions for these options. For knock-in, knock-out, and other barrier options, however, a boundary condition must be imposed at each barrier F_{bar}. Unless this barrier occurs at a natural boundary, where $C(F_{bar}) = 0$, the imposed boundary condition gives rise to a two-dimensional boundary layer which must be accounted for in pricing[11]. Thus, *barrier options should not be priced directly from the effective forward equations* without including an extra term that accounts for the boundary layer[11].

2.1. Effective media theory. We would like to apply the analysis of Appendix A to eqs. 2.2a-2.2c to obtain the implied volatilities for European options. We are seemingly precluded by the time-dependence of the coefficients $\sigma(T)$, $b(T)$, and $c(T)$. In [15] we sidestep this issue by using "effective media" arguments. These arguments show that for any given $T_{ex} > 0$, we can obtain the *same* density function $Q(T_{ex}, F)$, and thus the *same* European option prices and the *same* implied volatilities, by replacing $\sigma(T)$, $b(T)$, $c(T)$ and α with the appropriate constant coefficients $\Delta, \bar{b}, \bar{c}, \bar{\alpha}$. This result is not exact, but it *is* accurate through $O(\varepsilon^2)$, the same accuracy as the explicit implied volatility formulas.

To express the results succinctly, choose any exercise date T_{ex} and define

$$\tau(T) = \int_0^T \sigma^2\left(T'\right)dT', \tag{2.3a}$$

$$\tau_{ex} \equiv \tau(T_{ex}) = \int_0^{T_{ex}} \sigma^2\left(T'\right)dT'. \tag{2.3b}$$

In [15] we find that

$$Q(T_{ex}, F) \equiv \hat{Q}(T_{ex}, F) \qquad \text{for all } F \tag{2.4}$$

to within $O(\varepsilon^2)$, where $\hat{Q}(T, F)$ is the solution of

$$\hat{Q}_T = \tfrac{1}{2}\varepsilon^2\Delta^2\bar{\alpha}^2\left[\left(1 + 2\varepsilon\bar{b}z/\bar{\alpha} + \varepsilon^2\bar{c}z^2/\bar{\alpha}^2\right)C^2(F)\hat{Q}\right]_{FF} \qquad \text{for } 0 < T < T_{ex}, \tag{2.5a}$$

$$\hat{Q} = \delta(F - f) \qquad \text{as } T \to 0, \tag{2.5b}$$

with

$$z = \frac{1}{\varepsilon}\int_f^F \frac{dF'}{C(F')}. \tag{2.5c}$$

Here, the "effective" constant coefficients are defined by

$$\Delta^2 = \frac{1}{T_{ex}} \int_0^{T_{ex}} \sigma^2(T')dT' = \frac{\tau_{ex}}{T_{ex}}, \tag{2.6a}$$

$$\bar{b} = \frac{2}{\tau_{ex}^2} \int_0^{T_{ex}} \sigma^2(T)\,\tau(T)\,b(T)dT, \tag{2.6b}$$

$$\bar{c} = \frac{3}{\tau_{ex}^3} \int_0^{T_{ex}} \sigma^2(T)\,\tau^2(T)\,c(T)dT \tag{2.6c}$$

$$+\frac{18}{\tau_{ex}^3} \int_0^{T_{ex}} \sigma^2(T)\,b(T) \int_0^{T} \sigma^2(T_1)\,\tau(T_1)\,b(T_1)dT_1\,dT - 3\bar{b}^2, \tag{2.6d}$$

and

$$\bar{\alpha} = \alpha e^{\frac{1}{2}\varepsilon^2 \bar{A}}, \tag{2.6e}$$

with

$$\bar{A} = \frac{1}{2\tau_{ex}} \int_0^{T_{ex}} \sigma^2(T)\,\tau(T)\left[c(T) - \bar{c}\right]dT. \tag{2.6f}$$

Therefore, *for* all *the models listed in Appendix B, the implied normal volatilities* $\sigma_N(T_{ex}, K)$ *are given by the* same *equations, eqs.* 1.13a-1.13d, *through* $O(\varepsilon^2)$, *provided we set*

$$\alpha = \Delta\bar{\alpha}, \qquad \rho = \bar{b}/\sqrt{\bar{c}}, \qquad \nu = \Delta\sqrt{\bar{c}}. \tag{2.7}$$

This is the basis for our claim that these smile formulas are "universal." Note that the constants Δ, $\bar{\alpha}$, $\bar{b}$, $\bar{c}$ are different for different T_{ex}. This is to be expected since these constants represent an average of the time dependent coefficients over $0 < T < T_{ex}$.

The constant coefficients Δ, $\bar{\alpha}$, $\bar{b}$, $\bar{c}$ can be expressed more simply by changing the time variable from T to τ. Then,

$$\Delta^2 = \frac{\tau_{ex}}{T_{ex}} \tag{2.8a}$$

$$\bar{b} = \frac{2}{\tau_{ex}^2} \int_0^{\tau_{ex}} \tau b(\tau)d\tau, \tag{2.8b}$$

$$\bar{c} = \frac{3}{\tau_{ex}^3} \int_0^{\tau_{ex}} \tau^2 c(\tau)d\tau + \frac{18}{\tau_{ex}^3} \int_0^{\tau_{ex}} b(\tau) \int_0^{\tau} \tau_1 b(\tau_1)d\tau_1 d\tau - 3\bar{b}^2, \tag{2.8c}$$

and

$$\bar{\alpha} = \alpha e^{\frac{1}{2}\varepsilon^2 \bar{A}}, \qquad \bar{A} = \frac{1}{2\tau_{ex}} \int_0^{\tau_{ex}} \tau\left[c(\tau) - \bar{c}\right]d\tau. \tag{2.8d}$$

2.2. Numerical approach. An alternative approach is to numerically solve the effective forward equation

$$Q_T = \tfrac{1}{2}\varepsilon^2\alpha^2\sigma^2(T)\left[\left(1 + 2\varepsilon b(T)\,z/\alpha + \varepsilon^2 c(T)\,z^2/\alpha^2\right)C^2(F)Q\right]_{FF}, \tag{2.9a}$$

$$z \equiv \frac{1}{\varepsilon} \int_f^F \frac{dF'}{C(F')}, \tag{2.9b}$$

with the initial condition

$$Q = \delta(F - f) \qquad \text{as } T \to 0, \tag{2.9c}$$

and then integrate to find the European option values. Solving the problem numerically requires a finite domain, $F_{\min} < F < F_{\max}$. It is natural to choose diffusive barriers, points where $C(F) = 0$, for $F_{\min}$ and $F_{\max}$. However, if $C(F) > 0$ for all F, or if $C(F)$ has only a single zero, then we need to select artificial boundaries $F_{\min}$ small enough and/or $F_{\max}$ large enough, so that only an insignificant amount of probability can reach the boundaries.[4]

Boundary conditions for effective forward equations have been investigated in [4]. There it is found that absorbing boundary conditions must be used,

$$C^2(F)Q(T,F) = 0 \text{ as } F \to F_{\min}, \tag{2.10a}$$

$$C^2(F)Q(T,F) = 0 \text{ as } F \to F_{\max}, \tag{2.10b}$$

for the forward price F to be a Martingale.

We do not allow $\tilde{F}(T)$ to leave the domain $F_{\min} \le \tilde{F}(T) \le F_{\max}$, so any finite flux of probability from the interior of the interval $(F_{\min}, F_{max})$ to the barriers at $F = F_{\min}$ and $F = F_{\max}$, must accumulate as delta functions at the boundaries,

$$Q(T,F) = Q^L(T)\delta(F - F_{\min}) \qquad \text{at } F = F_{\min}, \tag{2.11a}$$

$$Q(T,F) = Q^R(T)\delta(F - F_{\max}) \qquad \text{at } F = F_{\max}. \tag{2.11b}$$

Conservation requires that the probabilities $Q^L(T)$ and $Q^R(T)$ increase according to the fluxes into the boundaries, so

$$\frac{dQ^L}{dT} = \lim_{F \to F_{\min}} \tfrac{1}{2}\varepsilon^2\alpha^2\sigma^2(T)\left[\left(1 + 2\varepsilon bz/\alpha + \varepsilon^2 cz^2/\alpha^2\right)C^2(F)Q\right]_F, \tag{2.12a}$$

$$\frac{dQ^L}{dT} = -\lim_{F \to F_{\max}} \tfrac{1}{2}\varepsilon^2\alpha^2\sigma^2(T)\left[\left(1 + 2\varepsilon bz/\alpha + \varepsilon^2 cz^2/\alpha^2\right)C^2(F)Q\right]_F. \tag{2.12b}$$

See the discussion in [4] for a physical interpretation of these delta functions as unresolved boundary layers.

The effective forward equation 2.9a-2.9c and boundary conditions 2.10a, 2.10b form a well-posed problem for obtaining $Q(T,F)$ for $F_{\min} < F < F_{\max}$, and 2.12a, 2.12b allow us to calculate the probability Q^L, Q^R tied up in the delta functions. The European option prices are then obtained by numerical integration,

$$V_{call}(T_{ex}, K) = \int_K^{F_{\max}} (F - K)Q(T_{ex}, F)dF + (F_{\max} - K)\,Q^R(T_{ex}), \tag{2.13a}$$

$$V_{put}(T_{ex}, K) = \int_{F_{\min}}^{K} (K - F)Q(T_{ex}, F)dF + (K - F_{\min})\,Q^L(T_{ex}), \tag{2.13b}$$

for $F_{\min} < K < F_{\max}$. For strikes K outside the interval $(F_{\min}, F_{\max})$, the options' time value is zero.

Together, the forward equation 2.9a-2.9c, the absorbing boundary conditions 2.10a, 2.10b, and the boundary accumulation equations 2.12a, 2.12b can be considered as a smile model for European options. It

[4]If $C(F_{\min}) = 0$, then the contribution from the two dimensional boundary layer is smaller than $O(\varepsilon^2)$, and hence negligible[11]. Same if $C(F_{\max}) = 0$. Otherwise, if $F_{\min}$ is small enough and $F_{\max}$ is large enough so that there is only a neglibible probability of reaching these boundaries, then the boundary layer contribution is also negligible.

appears to be a universal smile model, since the effective forward equation matches every stochastic volatility model that we have investigated, at least through $O\left(\varepsilon^2\right)$. In addition, it is easily seen that

$$\frac{d}{dT}\left\{Q^L(T_{ex}) + \int_{F_{\min}}^{F_{\max}} Q(T_{ex}, F)dF + Q^R(T_{ex})\right\} = 0, \tag{2.14a}$$

$$\frac{d}{dT}\left\{F_{\min}Q^L(T_{ex}) + \int_{F_{\min}}^{F_{\max}} FQ(T_{ex}, F)dF + F_{\max}Q^R(T_{ex})\right\} = 0, \tag{2.14b}$$

so the effective forward equation and boundary conditions conserve probability *exactly*, and $\tilde{F}(T)$ is *exactly* a Martingale. Provided that $\alpha^2\left(1 + 2\varepsilon bz/\alpha + \varepsilon^2\tilde{c}z^2/\alpha^2\right)C^2(F) > 0$ for all $F_{\min} < F < F_{\max}$, the maximum principle for parabolic equations[12] guarantees that $Q(T, F) > 0$ for $F_{\min} < F < F_{\max}$, and that $Q^L(T)$ and $Q^R(T)$ are positive for $T > 0$. Thus, the option prices defined by 2.9a-2.13b are arbitrage free[13], [14].

Conservative finite difference schemes have been developed to solve these equations in [32]. Since the effective forward equation has only one space dimension, solving the PDE is very quick. Moreover, solving the PDE once for $0 < T < T_{ex}$ yields $Q(T_{ex}, F)$ for all F, and thus the option prices and implied normal volatilities $\sigma_N(T_{ex}, K)$ for all strikes K. The conservative properties of the schemes ensure that the numerical solution also preserves probability and the Martingale nature of $\tilde{F}(T)$ exactly. For fine enough grids, the numerical solution for $Q\left(T, F\right)$ will also be non-negative, so the numerical solution itself constitutes an arbitrage free model[13], [14].

Finally, one can use numerical methods to solve the stationary equations obtained from the effective media analysis. These numerical solutions would be within $O\left(\varepsilon^2\right)$ of the solutions of the time-dependent PDE, and directly comparable to the explicit implied volatility smiles in 1.13a-1.13d.

3. Conclusions. Since so many different stochastic volatility models lead to the same effective forward equation, it would seem to be more efficient to directly use the effective forward equation itself,

$$Q_T = \tfrac{1}{2}\varepsilon^2\alpha^2\sigma^2\left(T\right)\left[\left(1 + 2\varepsilon b\left(T\right)z/\alpha + \varepsilon^2 c\left(T\right)z^2/\alpha^2\right)C^2(F)Q\right]_{FF}, \tag{3.1a}$$

$$z \equiv \frac{1}{\varepsilon}\int_f^F \frac{dF'}{C(F')}, \tag{3.1b}$$

as our smile model, rather than work with a more fundamental model like the Heston or mean reverting SABR model. For example, we could set α to be constant, and assume that $\sigma\left(T\right)$, $b\left(T\right)$, and $c\left(T\right)$ are piecewise constant, and then calibrate these coefficients to the smiles at each different exercise date. This would eliminate the tedious step of translating these coefficients back into the parameters of the more fundamental Heston or mean reverting SABR model.

At first glance, it is surprising that so many models lead to the same effective forward equation. We believe this occurs because we are analyzing essentially the same asymptotic regime of low to moderate vol of vol for each model. The factor

$$\alpha^2\sigma^2\left(T\right)\left(1 + 2\varepsilon b\left(T\right)z/\alpha + \varepsilon^2 c\left(T\right)z^2/\alpha^2\right) \tag{3.2}$$

may just be the first few terms in a Taylor series expansion of some function $E^2(T, \varepsilon z)$,

$$E^2(T, \varepsilon z) = \alpha^2\sigma^2\left(T\right)\left\{1 + 2\varepsilon b\left(T\right)z/\alpha + \varepsilon^2 c\left(T\right)z^2/\alpha^2 + \cdots\right., \tag{3.3}$$

which accounts for the systematic gains/losses due to the option's gamma with respect to both the forward asset price and the volatility. Possibly with more powerful techniques, we would be able to pin down $E^2(T, \varepsilon z)$ more exactly.

Here we note that if the factor $\alpha^2\sigma^2(T)\left(1+2\varepsilon b(T)z/\alpha+\varepsilon^2 c(T)z^2/\alpha^2\right)$ was replaced by any smooth positive function $E^2(T,\varepsilon z)$ in 2.9a-2.13b, and if $E^2(T,\varepsilon z)$ was bounded away from zero, $E^2(T,\varepsilon z)\geq\delta$ for some $\delta>0$, then the European option prices would remain arbitrage free. Indeed, the maximum principle[12] would still ensure that $Q(T,F)>0$ for $F_{\min}<F<F_{\max}$, and that $Q^L(T)$ and $Q^R(T)$ are positive, and the conservation laws 2.14a, 2.14b would still remain true, ensuring that probability is conserved exactly and that $\tilde{F}(T)$ is a Martingale. See [13], [14].

Appendix A. Asymptotic analysis of the effective forward equation.

Here we analyze the effective forward equation

$$Q_T=\tfrac{1}{2}\varepsilon^2\alpha^2\left[E_c^2(\varepsilon z)C^2(F)Q\right]_{FF} \tag{A.1}$$

to obtain explicit formulas for the implied normal volatilities of European options in the asymptotic limit $\varepsilon\ll 1$. These results will be accurate to within $O\left(\varepsilon^2\right)$. We first obtain the formulas for a general $E_c(\varepsilon z)$, assuming only that $E_c(0)=1$, that $E_c(\varepsilon z)$ is smooth, and that it is bounded away from zero,

$$E_c(\varepsilon z)\geq\delta \qquad \text{for some } \delta>0. \tag{A.2}$$

We then specialize the results to

$$E_c(\varepsilon z)\equiv\min\left\{\sqrt{1+2\varepsilon\rho\nu z/\alpha+\varepsilon^2\nu^2z^2/\alpha^2},E_{\max}\right\}, \tag{A.3}$$

to obtain the implied vols for the SABR model with a capped SVE quoted in §1.5. Letting the cap $E_{\max}\to\infty$ then yields the results for the original, uncapped, SABR model quoted in §1.2.

Spatial boundaries have a negligible impact on the implied volatilities, provided that neither today's forward f nor the strike K is within $O(\varepsilon)$ of a boundary, so the correct implied volatility formulas could be derived using free space arguments. Instead we analyze the effective forward equation over a finite domain $F_{\min}<F<F_{\max}$, and show that the effect of boundaries is transcendentally small, and thus negligible to all orders, when both f and K are away from the boundaries.

Define

$$D^2(F)=E_c^2(\varepsilon z)C^2(F), \tag{A.4a}$$

with

$$z(F)=\frac{1}{\varepsilon}\int_f^F\frac{dF'}{C(F')}, \tag{A.4b}$$

and let

$$\tau=\alpha^2T, \qquad \tau_{ex}=\alpha^2T_{ex}. \tag{A.5}$$

The effective forward equation is then

$$Q_\tau=\tfrac{1}{2}\varepsilon^2\left[D^2(F)Q\right]_{FF} \qquad \text{for } F_{\min}<F<F_{\max}, \tag{A.6a}$$

with the initial conditions

$$Q(\tau,F)=\delta(F-f) \qquad \text{as } \tau\to 0. \tag{A.6b}$$

In [4] it is shown that the Martingale property of $\tilde{F}(T)$ requires that absorbing boundary conditions be used, so the boundary conditions are

$$D^2(F)Q\to 0 \qquad \text{as } F\to F_{\min} \text{ and as } F\to F_{\max}. \tag{A.6c}$$

We do not allow the forward $\tilde{F}(T)$ to leave the domain $F_{\min} \leq \tilde{F}(T) \leq F_{\max}$, so any leakage of probability out of the interior of $(F_{\min}, F_{max})$ must accumulate as delta functions at the boundaries,

$$Q(T,F) = Q^L(T)\delta(F - F_{\min}) \qquad \text{at } F = F_{\min}, \tag{A.7a}$$

$$Q(T,F) = Q^R(T)\delta(F - F_{\max}) \qquad \text{at } F = F_{\max}. \tag{A.7b}$$

See the discussion in [4] for a physical interpretation of these delta functions as unresolved boundary layers. Conservation requries that the flux of probability to the boundaries must balance the increase in probability in the delta functions, so

$$\frac{dQ^L}{d\tau} = \lim_{F \to F_{\min}} \tfrac{1}{2}\varepsilon^2 \left[D^2(F)Q\right]_F, \tag{A.7c}$$

$$\frac{dQ^R}{d\tau} = -\lim_{F \to F_{\max}} \tfrac{1}{2}\varepsilon^2 \left[D^2(F)Q\right]_F. \tag{A.7d}$$

After solving the forward problem A.6a-A.6c for $Q(\tau, F)$, and integrating A.7c, A.7d for $Q^L(\tau)$ and $Q^R(\tau)$, the value of a European option with expiry τ_{ex} and strike K is

$$V_{call}(\tau_{ex}, K) = \int_K^{F_{\max}} (F - K)Q(\tau_{ex}, F)\,dF + (F_{\max} - K)\,Q^R(\tau_{ex}), \tag{A.8a}$$

$$V_{put}(\tau_{ex}, K) = (K - F_{\min})\,Q^L(\tau_{ex}) + \int_{F_{\min}}^K (K - F)Q(\tau_{ex}, F)\,dF. \tag{A.8b}$$

We will find that the contributions from the boundary terms Q^L, Q^R are transcendentally small.

Using A.6a we can write

$$\begin{aligned} Q(\tau_{ex}, F) &= \delta(F - f) + \int_0^{\tau_{ex}} Q_\tau(\tau, F)d\tau \\ &= \delta(F - f) + \tfrac{1}{2}\varepsilon^2 \int_0^{\tau_{ex}} \left[D^2(F)Q(\tau, F)\right]_{FF} d\tau, \end{aligned} \tag{A.9a}$$

$$Q^R(\tau_{ex}) = -\lim_{F \to F_{\max}} \tfrac{1}{2}\varepsilon^2 \int_0^{\tau_{ex}} \left[D^2(F)Q(\tau, F)\right]_F d\tau. \tag{A.9b}$$

Substituting this into A.8a yields

$$\begin{aligned} V_{call}(\tau_{ex}, K) = [f - K]^+ + \tfrac{1}{2}\varepsilon^2 \int_0^{\tau_{ex}} \Bigg\{ &\int_K^{F_{\max}} (F - K)\left[D^2(F)Q(\tau, F)\right]_{FF} dF \\ &- \lim_{F \to F_{\max}} (F_{\max} - K)\left[D^2(F)Q(\tau, F)\right]_F \Bigg\} d\tau, \end{aligned} \tag{A.10}$$

and integrating by parts twice yields

$$V_{call}(\tau_{ex}, K) = [f - K]^+ + \tfrac{1}{2}\varepsilon^2 D^2(K) \int_0^{\tau_{ex}} Q(\tau, K)d\tau. \tag{A.11a}$$

A similar argument gives

$$V_{put}(\tau_{ex}, K) = [K - f]^+ + \tfrac{1}{2}\varepsilon^2 D^2(K) \int_0^{\tau_{ex}} Q(\tau, K)d\tau. \tag{A.11b}$$

These expressions are helpful, since the accuracy of the option price depends only on the accuracy of the formulas for $Q(\tau, K)$ at strike K. Stripping off the intrinsic values $[f-K]^{+}$ and $[K-f]^{+}$ of the call and put prices, we work directly with the European options' time value:

$$\begin{aligned} V_{tv}(\tau_{ex}, K) &\equiv V_{call}(\tau_{ex}, K) - [f-K]^{+} \equiv V_{put}(\tau_{ex}, K) - [K-f]^{+} \\ &= \tfrac{1}{2}\varepsilon^2 D^2(K) \int_0^{\tau_{ex}} Q(\tau, K) d\tau. \end{aligned} \tag{A.12}$$

A.1. Near identity transform. We use a variant of the near-identity transform technique [23], [1], where the forward problem is successively transformed until it is solvable by inspection. Define $\tilde{Q}$ by

$$\tilde{Q}(\tau, F) = \frac{D^2(F)}{D^2(f)} Q(\tau, F). \tag{A.13}$$

Then the forward problem A.6a-A.6c becomes

$$\tilde{Q}_\tau = \tfrac{1}{2}\varepsilon^2 D^2(F) \tilde{Q}_{FF} \qquad \text{for } F_{\min} < F < F_{\max}, \tag{A.14a}$$

$$\tilde{Q} \to 0 \qquad \text{as } F \to F_{\min} \text{ and } F \to F_{\min}, \tag{A.14b}$$

$$\tilde{Q} \to \delta(F-f) \qquad \text{as } \tau \to 0, \tag{A.14c}$$

and the time-value becomes

$$V_{tv}(\tau_{ex}, K) = \tfrac{1}{2}\varepsilon^2 D^2(f) \int_0^{\tau_{ex}} \tilde{Q}(\tau, K) d\tau. \tag{A.15}$$

Since this is a diffusion problem, we anticipate thin, $O(\varepsilon)$ width, boundary layers adjacent to $F_{\min}$ and $F_{\max}$. We require the forward f and strike K to be in the "outer region" outside these boundary layers,

$$(f - F_{\min})/\varepsilon \gg 1, \qquad (F_{\max} - f)/\varepsilon \gg 1, \tag{A.16a}$$

$$(K - F_{\min})/\varepsilon \gg 1, \qquad (F_{\max} - K)/\varepsilon \gg 1. \tag{A.16b}$$

Define

$$z = \frac{1}{\varepsilon} \int_f^F \frac{dF'}{C(F')}, \qquad Z = \frac{1}{\varepsilon} \int_f^K \frac{dF'}{C(F')}, \tag{A.17a}$$

and for clarity, let

$$B(\varepsilon z) \equiv C(F). \tag{A.17b}$$

Since $D(F) = E(\varepsilon z) C(F)$, we define a new independent variable,

$$x = \frac{1}{\varepsilon} \int_f^F \frac{dF'}{D(F')} = \frac{1}{\varepsilon} \int_f^F \frac{dF'}{E_c(\varepsilon z(F'))C(F')} = \int_0^{z(F)} \frac{dz'}{E_c(\varepsilon z')} = I(\varepsilon z(F)), \tag{A.18a}$$

where

$$I(\varepsilon z) \equiv \int_0^z \frac{dz'}{E_c(\varepsilon z')} = \frac{1}{\varepsilon} \int_0^{\varepsilon z} \frac{dq'}{E_c(q')}, \tag{A.18b}$$

and let

$$X = x(K) = I(\varepsilon z(K)) = I(\varepsilon Z). \tag{A.18c}$$

For clarity, we also define

$$M(\varepsilon x) \equiv D(F) = E_c(\varepsilon z)B(\varepsilon z), \tag{A.18d}$$

where $\varepsilon z\,(\varepsilon x)$ is defined implicitly by $\varepsilon x = \varepsilon I\,(\varepsilon z)$.

The change of variables yields

$$\frac{\partial}{\partial F} \to \frac{1}{\varepsilon M(\varepsilon x)}\frac{\partial}{\partial x}, \tag{A.19a}$$

$$\frac{\partial^2}{\partial F^2} \to \frac{1}{\varepsilon^2 M^2(\varepsilon x)}\left\{\frac{\partial^2}{\partial x^2} - \varepsilon\frac{M'(\varepsilon x)}{M(\varepsilon x)}\frac{\partial}{\partial x}\right\}, \tag{A.19b}$$

and

$$\delta(F-f) = \frac{\delta(x)}{\varepsilon M(0)}. \tag{A.19c}$$

In terms of the new variables, the forward problem is

$$\tilde{Q}_\tau = \tfrac{1}{2}\tilde{Q}_{xx} - \tfrac{1}{2}\varepsilon\frac{M'(\varepsilon x)}{M(\varepsilon x)}\tilde{Q}_x \qquad \text{for } \tau > 0, \tag{A.20a}$$

$$\tilde{Q} \to \frac{\delta(x)}{\varepsilon M(0)} \qquad \text{as } \tau \to 0, \tag{A.20b}$$

and the option price is

$$V_{tv}(\tau_{ex}, K) = \tfrac{1}{2}\varepsilon^2 M^2(0)\int_0^{\tau_{ex}} \tilde{Q}(\tau, X)d\tau. \tag{A.21}$$

Finally, define $H(\tau, x)$ by

$$\tilde{Q}(\tau, x) = \frac{M^{1/2}(\varepsilon x)}{\varepsilon M^{3/2}(0)}H(\tau, x). \tag{A.22}$$

Then the effective forward problem becomes

$$H_\tau = \tfrac{1}{2}H_{xx} + \varepsilon^2\kappa(\varepsilon x)H, \qquad \text{for } \tau > 0.\backslash, \tag{A.23a}$$

$$H\,(\tau, x) \to \delta(x) \qquad \text{as } \tau \to 0, \tag{A.23b}$$

where

$$\kappa(\varepsilon x) = \tfrac{1}{4}\frac{M''(\varepsilon x)}{M(\varepsilon x)} - \tfrac{3}{8}\left(\frac{M'(\varepsilon x)}{M(\varepsilon x)}\right)^2. \tag{A.24}$$

Once we have solved for $H(\tau, x)$, the option's time value is

$$V_{tv}\,(\tau_{ex}, K) = \tfrac{1}{2}\varepsilon\sqrt{M(0)M(\varepsilon X)}\int_0^{\tau_{ex}} H(\tau, X)d\tau. \tag{A.25}$$

Since

$$\frac{dz}{dF} = \frac{1}{\varepsilon B(\varepsilon z)}, \qquad \frac{dx}{dz} = \frac{1}{E_c(\varepsilon z)}, \qquad \frac{dx}{dF} = \frac{1}{\varepsilon M\,(\varepsilon x)}, \tag{A.26}$$

we have

$$\kappa(\varepsilon x) = E^2(\varepsilon z)\Delta(\varepsilon z) + \tfrac{1}{4}E(\varepsilon z)E''(\varepsilon z) - \tfrac{1}{8}\left[E'(\varepsilon z)\right]^2, \tag{A.27a}$$

with

$$\Delta(\varepsilon z) \equiv \tfrac{1}{4}\frac{B''(\varepsilon z)}{B(\varepsilon z)} - \tfrac{3}{8}\left(\frac{B'(\varepsilon z)}{B(\varepsilon z)}\right)^2 = \tfrac{1}{4}C(F)C''(F) - \tfrac{1}{8}\left[C'(F)\right]^2. \tag{A.27b}$$

Now $\varepsilon^2\kappa(\varepsilon x)$, and thus $\varepsilon^2E^2(\varepsilon z)\Delta(\varepsilon z)$, enter equation A.23a at $O\left(\varepsilon^2\right)$. We can replace $\varepsilon^2\Delta(\varepsilon z)$ by $\varepsilon^2\Delta(\varepsilon z_0)$ with z_0 constant, while committing only $O\left(\varepsilon^3\right)$ errors. These errors are negligible since we are working through $O\left(\varepsilon^2\right)$. So we set

$$\Delta(\varepsilon z) \to \Delta_0 \equiv \Delta(0) = \tfrac{1}{4}C(f_0)C''(f_0) - \tfrac{1}{8}\left[C'(f_0)\right]^2. \tag{A.28}$$

Note that Δ is exactly constant for normal backbones, $C(F) = 1$, and for shifted log normal backbones, $C(F) \equiv F + o$, so there is no approximation for these key cases.

A.2. Asymptotic solution for H. To solve the effective forward equation A.23a, A.23b with

$$\kappa(\varepsilon x) = E^2(\varepsilon z)\Delta_0 + \tfrac{1}{4}E(\varepsilon z)E''(\varepsilon z) - \tfrac{1}{8}\left[E'(\varepsilon z)\right]^2, \tag{A.29}$$

we look for a solution of the form

$$H(\tau, x) = \frac{1}{\sqrt{2\pi\tau}}e^{-x^2/2\tau}e^{\varepsilon^2 S(\tau,\varepsilon x)}. \tag{A.30}$$

Let

$$\chi = \varepsilon x. \tag{A.31}$$

Substituting H into A.23a, A.23b yields

$$S_\tau + \frac{\chi}{\tau}S_\chi = \kappa(\chi) + \tfrac{1}{2}\varepsilon^2 S_{\chi\chi} + \tfrac{1}{2}\varepsilon^4 S_\chi^2, \tag{A.32a}$$

with the initial condition

$$S(0, \chi) = 0. \tag{A.32b}$$

Substituting the expansion

$$S(\tau, \chi) = \tau s^{(0)}(\chi) + \varepsilon^2\tau^2 s^{(1)}(\chi) + \cdots \tag{A.33a}$$

into eq. A.32a leads to the hierarchy of equations:

$$\chi s_\chi^{(0)} + s^{(0)} = \kappa(\chi), \tag{A.33b}$$

$$\chi s_\chi^{(1)} + 2s^{(1)} = \tfrac{1}{2}s_{\chi\chi}^{(0)}, \tag{A.33c}$$

Solving A.33b yields

$$H(\tau, x) = \frac{1}{\sqrt{2\pi\tau}}e^{-x^2/2\tau}e^{\varepsilon^2\tau s^{(0)}(\varepsilon x) + \cdots}, \tag{A.34a}$$

with

$$s^{(0)} = \frac{1}{\chi}\int_0^\chi \kappa(\chi')d\chi' = \frac{1}{x}\int_0^x \kappa(\varepsilon x')dx', \tag{A.34b}$$

as can be verified by direct substitution.

A.3. Option prices. With this $H(\tau, x)$, equation A.25 for the option value becomes

$$V_{tv}(\tau_{ex}, K) = \tfrac{1}{2}\varepsilon\sqrt{M(0)M(\varepsilon X)} \int_0^{\tau_{ex}} e^{-X^2/2\tau} e^{\varepsilon^2 \tau s^{(0)}(\varepsilon X) + \cdots} \frac{d\tau}{\sqrt{2\pi\tau}}, \tag{A.35}$$

where $X = x(K)$. We change integration variables to $q = X^2/2\tau$. Since

$$\frac{d\tau}{\sqrt{2\pi\tau}} = -\frac{|X|}{2\sqrt{\pi}} \frac{dq}{q^{3/2}}, \tag{A.36}$$

we obtain

$$\begin{aligned} V_{tv}(\tau_{ex}, K) &= \frac{\varepsilon\,|X|}{4\sqrt{\pi}} \sqrt{M(0)M(\varepsilon X)} \int_{X^2/2\tau_{ex}}^{\infty} \frac{e^{-q}}{q^{3/2}} e^{\varepsilon^2 X^2 s^{(0)}(\varepsilon X)/2q + \cdots} dq \\ &= \frac{\varepsilon\,|X|}{4\sqrt{\pi}} \sqrt{M(0)M(\varepsilon X)} \int_{X^2/2\tau_{ex}}^{\infty} \frac{e^{-q}}{q^{3/2}} \left\{1 + \frac{\varepsilon^2}{2q} X^2 s^{(0)}\left(\varepsilon X\right) + \cdots \right\} dq. \end{aligned} \tag{A.37}$$

But

$$\int_a^{\infty} \frac{e^{-q}}{q^{5/2}} dq = \tfrac{2}{3}\frac{e^{-a}}{a^{3/2}} - \tfrac{2}{3}\int_a^{\infty} \frac{e^{-q}}{q^{3/2}} dq, \tag{A.38}$$

so

$$\begin{aligned} V_{tv}(\tau_{ex}, K) = \frac{\varepsilon\,|X|}{4\sqrt{\pi}} \sqrt{M(0)M(\varepsilon X)} \cdot \Bigg\{ \left(1 - \tfrac{1}{3}\varepsilon^2 X^2 s^{(0)}\left(\varepsilon X\right) + \cdots \right) \int_a^{\infty} \frac{e^{-q}}{q^{3/2}} dq \\ + \tfrac{1}{3}\varepsilon^2 X^2 s^{(0)}\left(\varepsilon X\right) \frac{e^{-a}}{a^{3/2}} + \cdots \Bigg\} \end{aligned} \tag{A.39}$$

with $a = X^2/2\tau_{ex}$. Since

$$\frac{\partial}{\partial a} \int_a^{\infty} \frac{e^{-q}}{q^{3/2}} dq = -\frac{e^{-a}}{a^{3/2}}, \tag{A.40}$$

this is

$$V_{tv}(\tau_{ex}, K) = \frac{\varepsilon\,|X|}{4\sqrt{\pi}} \sqrt{M(0)M(\varepsilon X)} \left(1 - \tfrac{1}{3}\varepsilon^2 X^2 s^{(0)}\left(\varepsilon X\right) + \cdots \right) \int_{a^*}^{\infty} \frac{e^{-q}}{q^{3/2}} dq \tag{A.41a}$$

through $O(\varepsilon^2)$, where

$$a^* = \frac{X^2}{2\tau_{ex}} \left\{1 - \tfrac{2}{3}\varepsilon^2 s^{(0)}\left(\varepsilon X\right) \tau_{ex} + \cdots \right\}. \tag{A.41b}$$

Remarkably,

$$K - f = \varepsilon X \sqrt{M(0)M(\varepsilon X)} \left(1 - \tfrac{1}{3}\varepsilon^2 X^2 s^{(0)}\left(\varepsilon X\right) + \cdots \right) \tag{A.42}$$

through $O(\varepsilon^2)$. To show this, recall that

$$\frac{dF}{dx} = \varepsilon D(F) = \varepsilon M(\varepsilon x). \tag{A.43}$$

See eqs. A.18a-A.18d. Hence,

$$\text{(A.44)} \qquad \frac{K-f}{\varepsilon X\sqrt{M(0)M(\varepsilon X)}} = \frac{1}{X}\int_0^X \frac{M(\varepsilon x)}{\sqrt{M(0)M(\varepsilon X)}}dx.$$

Expanding around the midpoint $\frac{1}{2}\varepsilon X$ leads to

$$\text{(A.45a)} \qquad \frac{K-f}{\varepsilon X\sqrt{M(0)M(\varepsilon X)}} = 1 - \tfrac{1}{24}\left(2\gamma_2 - 3\gamma_1^2\right)\varepsilon^2 X^2 + O(\varepsilon^4),$$

where

$$\text{(A.45b)} \qquad \gamma_1 = \frac{M'\left(\frac{1}{2}\varepsilon X\right)}{M\left(\frac{1}{2}\varepsilon X\right)}, \qquad \gamma_2 = \frac{M''\left(\frac{1}{2}\varepsilon X\right)}{M\left(\frac{1}{2}\varepsilon X\right)}.$$

Similarly, A.24 shows that

$$\text{(A.46)} \qquad s^{(0)}\left(\varepsilon X\right) = \frac{1}{X}\int_0^X \kappa(\varepsilon x')dx' = \kappa(\tfrac{1}{2}\varepsilon X) + O\left(\varepsilon^2\right) = \tfrac{1}{4}\gamma_2 - \tfrac{3}{8}\gamma_1^2 + O\left(\varepsilon^2\right),$$

so

$$\text{(A.47)} \qquad 1 - \tfrac{1}{3}\varepsilon^2 X^2 s^{(0)}\left(\varepsilon X\right) + \cdots = 1 - \tfrac{1}{24}(2\gamma_2 - 3\gamma_1^2)\varepsilon^2 X^2 + O\left(\varepsilon^4\right).$$

Clearly the expressions in A.45a and A.47 are equal up to $O\left(\varepsilon^4\right)$, establishing A.42, so the option price is

$$\text{(A.48a)} \qquad V_{tv}(\tau_{ex}, K) = \frac{|K-f|}{4\sqrt{\pi}}\int_{a^*}^{\infty}\frac{e^{-q}}{q^{3/2}}dq,$$

through $O\left(\varepsilon^2\right)$, with

$$\text{(A.48b)} \qquad a^* = \frac{X^2}{2\tau_{ex}}\left\{1 - \tfrac{2}{3}\varepsilon^2 s^{(0)}\left(\varepsilon X\right)\tau_{ex} + \cdots\right\}.$$

A.4. Implied normal vol. Equations A.48a, A.48b provide an explicit expression for the time value of European options. We can re-express this price more intuitively as an implied normal volatility. Under the normal (Bachelier) model,

$$\text{(A.49a)} \qquad dF = \sigma_N dW,$$

the time value of a European option option is

$$\text{(A.49b)} \qquad \begin{aligned} V_{tv}^N(\tau_{ex}, K) &= -|K-f|\,\mathcal{N}\left(-\frac{|K-f|}{\sigma_N T_{ex}^{1/2}}\right) + \sigma_N T_{ex}^{1/2}\mathcal{G}\left(-\frac{|K-f|}{\sigma_N T_{ex}^{1/2}}\right) \\ &= \frac{|K-f|}{4\sqrt{\pi}}\int_{(K-f)^2/2\sigma_N^2 T_{ex}}^{\infty}\frac{e^{-q}}{q^{3/2}}dq, \end{aligned}$$

where $\mathcal{N}(\cdot)$ is the cumulative normal distribution and $\mathcal{G}(\cdot)$ is the normal density. The option value in A.48a, A.48b matches the option value under the normal model, provided that we choose the volatility σ_N so that

$$\text{(A.50)} \qquad \frac{(K-f)^2}{2\sigma_N^2 T_{ex}} = a^* = \frac{X^2}{2\tau_{ex}}\left\{1 - \tfrac{2}{3}\varepsilon^2 s^{(0)}\left(\varepsilon X\right)\tau_{ex} + \cdots\right\}.$$

Recalling that $\tau_{ex} = \alpha^2 T_{ex}$, we solve A.50 for the implied normal volatilty σ_N, finding that

$$\sigma_N\left(T_{ex}, K\right) = \alpha\frac{K-f}{X}\left\{1+\varepsilon^2\theta T_{ex} - \cdots\right\} \tag{A.51a}$$

through $O(\varepsilon^2)$, with

$$\theta = \tfrac{1}{3}\alpha^2 s^{(0)}\left(\varepsilon X\right). \tag{A.51b}$$

At this order we cannot distinguish between $1+\varepsilon^2\theta T_{ex}$ and $1/[1-\varepsilon^2\theta T_{ex}]$; both are equally accurate through $O(\varepsilon^2)$. As a practical matter, we choose

$$\sigma_N\left(T_{ex}, K\right) = \alpha\frac{K-f}{X}\begin{cases} 1+\varepsilon^2\theta T_{ex} & \text{if } \theta \geq 0 \\ \dfrac{1}{1-\varepsilon^2\theta T_{ex}} & \text{if } \theta < 0 \end{cases}. \tag{A.52}$$

Under moderate conditions, when $\varepsilon^2\left|\theta\right|$ is small, the two forms are essentially equivalent. Under extreme conditions, however, when $\varepsilon^2\left|\theta\right|$ is not particularly small, eq. A.52 is safer than using either form exclusively. As a bonus, it also seems to be more accurate.

Although this expansion can be continued to higher order, the complexity and fragility of the resulting implied volatility formulas overwhelms any increase in accuracy

A.5. Explicit formulas. Adding the intrinsic and time values together yields the prices,

$$V_{call}(T_{ex}, K) = (f-K)\mathcal{N}\left(\frac{f-K}{\sigma_N T_{ex}^{1/2}}\right) + \sigma_N T_{ex}^{1/2}\mathcal{G}\left(\frac{f-K}{\sigma_N T_{ex}^{1/2}}\right), \tag{A.53a}$$

$$V_{put}(T_{ex}, K) = (K-f)\mathcal{N}\left(\frac{K-f}{\sigma_N T_{ex}^{1/2}}\right) + \sigma_N T_{ex}^{1/2}\mathcal{G}\left(\frac{K-f}{\sigma_N T_{ex}^{1/2}}\right), \tag{A.53b}$$

where implied normal volatility $\sigma_N\left(T_{ex}, K\right)$ is given by eq. A.52 above. We can write $\sigma_N\left(T_{ex}, K\right)$ in a more intuitive form by recalling that

$$Z = z\left(K\right) = \frac{1}{\varepsilon}\int_f^K \frac{dF}{C\left(F\right)} \tag{A.54a}$$

$$X = x\left(K\right) \equiv \int_0^Z \frac{dz'}{E_c(\varepsilon z')} \equiv I(\varepsilon Z). \tag{A.54b}$$

Then $\sigma_N\left(T_{ex}, K\right)$ can be written as

$$\sigma_N\left(T_{ex}, K\right) = \varepsilon\alpha \cdot \frac{1}{\dfrac{1}{K-f}\displaystyle\int_f^K \frac{dF}{C\left(F\right)}} \cdot \frac{1}{\dfrac{1}{Z}\displaystyle\int_0^Z \frac{dz'}{E_c(\varepsilon z')}} \cdot \begin{cases} 1+\varepsilon^2\theta T_{ex} & \text{if } \theta \geq 0 \\ \dfrac{1}{1-\varepsilon^2\theta T_{ex}} & \text{if } \theta < 0 \end{cases} \tag{A.55a}$$

through $O(\varepsilon^2)$, where the second order terms are given by

$$\theta = \tfrac{1}{3}\frac{\alpha^2}{X}\int_0^X \kappa(\varepsilon x)dx, \tag{A.55b}$$

with

$$\kappa(\varepsilon x) = E^2(\varepsilon z)\Delta_0 + \tfrac{1}{4}E(\varepsilon z)E''(\varepsilon z) - \tfrac{1}{8}\left[E'(\varepsilon z)\right]^2, \tag{A.55c}$$

$$\Delta_0 = \tfrac{1}{4}C(f)C''\left(f\right) - \tfrac{1}{8}\left[C'\left(f\right)\right]^2. \tag{A.55d}$$

A.5.1. SABR with a capped stochastic volatility effect. Let us now specialize to the SABR model with a capped stochastic volatility effect (SVE):

$$E_c(\varepsilon z) = \min\left\{\sqrt{1+2\varepsilon\rho\nu z/\alpha+\varepsilon^2\nu^2 z^2/\alpha^2}, E_{\max}\right\}. \tag{A.56}$$

The implied normal volatility can be expressed most cleanly in terms of $\zeta = \varepsilon\nu Z/\alpha$ and

$$Y \equiv \varepsilon\frac{\nu}{\alpha}I(\varepsilon Z) = \frac{\varepsilon\nu}{\alpha}\int_0^Z \frac{dz'}{E_c(\varepsilon z')} = \int_0^\zeta \frac{d\zeta'}{\min\left\{\sqrt{1+2\rho\zeta'+\left(\zeta'\right)^2}, E_{\max}\right\}} \tag{A.57}$$

It is

$$\sigma\left(T_{ex}, K\right) = \varepsilon\alpha\cdot\frac{(K-f)}{\displaystyle\int_f^K \frac{dF}{C\left(F\right)}}\cdot\frac{\zeta}{Y\left(\zeta\right)}\cdot\begin{cases} 1+\varepsilon^2\theta\left(\zeta\right)T_{ex} & \text{if } \theta\geq 0\\ \dfrac{1}{1-\varepsilon^2\theta\left(\zeta\right)T_{ex}} & \text{if } \theta<0\end{cases}, \tag{A.58a}$$

with

$$\zeta = \frac{\nu}{\alpha}\int_f^K \frac{dF}{C\left(F\right)}, \tag{A.58b}$$

$$Y\left(\zeta\right) = \begin{cases} Y_- - \left(\zeta_- - \zeta\right)/E_{\max} & \text{if } \zeta<\zeta_-,\\ \log\dfrac{\sqrt{1+2\rho\zeta+\zeta^2}+\rho+\zeta}{1+\rho} & \text{if } \zeta_-<\zeta<\zeta_+,\\ Y_+ + \left(\zeta-\zeta_+\right)/E_{\max} & \text{if } \zeta_+<\zeta,\end{cases} \tag{A.58c}$$

and

$$\zeta_\pm = -\rho\pm\sqrt{E_{\max}^2-1+\rho^2}, \qquad Y_\pm = \pm\log\frac{E_{\max}+\sqrt{E_{\max}^2-1+\rho^2}}{1\pm\rho}. \tag{A.58d}$$

Here, $\zeta_\pm$ and $Y_\pm$ are the values of ζ, $Y\left(\zeta\right)$ at which $E(\zeta)$ hits $E_{\max}$, where

$$E(\zeta) = \sqrt{1+2\rho\zeta+\zeta^2}. \tag{A.58e}$$

For $\zeta<\zeta_-$, the higher order correction term θ is

$$\begin{aligned}\theta\left(\zeta\right) = {}& \frac{\nu^2}{24Y\left(\zeta\right)}\left\{-Y_- - 3\frac{\sqrt{E_{\max}^2-1+\rho^2}}{E_{\max}} - 3\rho\right\}\\ &+\frac{\Delta_0\alpha^2}{6Y\left(\zeta\right)}\left\{2E_{\max}\left(\zeta-\zeta_-\right)+\left(1-\rho^2\right)Y_- - E_{\max}\sqrt{E_{\max}^2-1+\rho^2}-\rho\right\}.\end{aligned} \tag{A.59a}$$

For $\zeta_- < \zeta < \zeta$, it is

$$\theta\left(\zeta\right) = \frac{\nu^2}{24}\left\{-1+3\frac{\rho+\zeta-\rho E\left(\zeta\right)}{Y\left(\zeta\right)E\left(\zeta\right)}\right\}+\frac{\Delta_0\alpha^2}{6}\left\{1-\rho^2+\frac{\left(\rho+\zeta\right)E\left(\zeta\right)-\rho}{Y\left(\zeta\right)}\right\}. \tag{A.59b}$$

Finally, for $\zeta > \zeta_+$, it is

$$\text{(A.59c)} \qquad \theta(\zeta) = \frac{\nu^2}{24Y(\zeta)} \left\{ -Y_+ + 3\frac{\sqrt{E_{\max}^2 - 1 + \rho^2}}{E_{\max}} - 3\rho \right\} + \frac{\Delta_0 \alpha^2}{6Y(\zeta)} \left\{ 2E_{\max}(\zeta - \zeta_+) + (1-\rho^2) Y_+ + E_{\max}\sqrt{E_{\max}^2 - 1 + \rho^2} - \rho \right\}.$$

In the above

$$\text{(A.60)} \qquad \Delta_0 = \tfrac{1}{4} C(f) C''(f) - \tfrac{1}{8} [C'(f)]^2.$$

A.5.2. Uncapped stochastic vol effect. For the standard SABR model, the SVE is unbounded, so,

$$\text{(A.61)} \qquad E_c(\varepsilon z) = E(\zeta) = \sqrt{1 + 2\varepsilon\rho\nu z/\alpha + \varepsilon^2 \nu^2 z^2/\alpha^2},$$

The implied normal vols for this model can be obtained from the preceding formulas by letting the cap $E_{\max}$ go to infinity. As $E_{\max} \to \infty$, $\zeta_- \to -\infty$ and $\zeta_+ \to \infty$. Thus, the implied normal volatilities $\sigma_N(T_{ex}, K)$ are again given by equation A.58a, with ζ, $Y(\zeta)$, and $E(\zeta)$ given by

$$\text{(A.62)} \qquad \zeta = \frac{\nu}{\alpha} \int_f^K \frac{dF}{C(F)}, \qquad Y(\zeta) = \log \frac{E(\zeta) + \rho + \zeta}{1 + \rho}, \qquad E(\zeta) = \sqrt{1 + 2\rho\zeta + \zeta^2},$$

and with θ given by A.59b above. These are equations 1.13a - 1.13d quoted in §1.2.

A.5.3. Offset CEV backbone. We can make these formulas more explicit by postulating a backbone function $C(F)$. The most commonly used backbone is the offset CEV function,

$$\text{(A.63a)} \qquad C(F) = (F + o)^\beta$$

where normally $0 \leq \beta \leq 1$. For this case, the local volatility factor is given by

$$\text{(A.63b)} \qquad \frac{(K - f)}{\displaystyle\int_f^K \frac{dF}{C(F)}} = (1 - \beta) \frac{K - f}{(K + o)^{1-\beta} - (f + o)^{1-\beta}},$$

and ζ and Δ_0 are given by

$$\text{(A.63c)} \qquad \zeta = \frac{\nu}{\alpha} \int_f^K \frac{dF}{C(F)} = \frac{\nu}{\alpha} \frac{(K + o)^{1-\beta} - (f + o)^{1-\beta}}{1 - \beta},$$

$$\text{(A.63d)} \qquad \Delta_0 = \tfrac{1}{4} C(f) C''(f) - \tfrac{1}{8} [C'(f)]^2 = -\tfrac{1}{8} \frac{\beta(2 - \beta)}{(f + o)^{2-2\beta}}.$$

Appendix B. Other models with effective forward equations.

B.1. SABR-type models. The dynamic SABR model is

$$d\tilde{F} = \varepsilon\sigma_0(T)\tilde{A}C(\tilde{F})d\tilde{W}_1, \tag{B.1a}$$

$$d\tilde{A} = \varepsilon\nu(T)\tilde{A}d\tilde{W}_2, \tag{B.1b}$$

$$d\tilde{W}_1 d\tilde{W}_2 = \rho(T)dT, \tag{B.1c}$$

and is identical to the original SABR model, except that $\sigma_0(T)$, $\rho(T)$, and $\nu(T)$ are time dependent. Often $\sigma_0(T)$, $\rho(T)$, and $\nu(T)$ are taken to be piecewise constant with the objective of calibrating the theoretical volatility smiles $\sigma_N(T_{ex}, K)$ to the market smiles at several different exercise dates.

In [16], it is shown that the marginal density $Q(T,F)$ for this model satisfies

$$Q_T = \tfrac{1}{2}\varepsilon^2\alpha^2\sigma^2(T)\left[\left(1 + 2\varepsilon b(T)z/\alpha + \varepsilon^2 c(T)z^2/\alpha^2\right)C^2(F)Q\right]_{FF}, \tag{B.2a}$$

$$Q \to \delta(F-f) \qquad \text{as } T \to 0 \tag{B.2b}$$

through $O\left(\varepsilon^2\right)$, where

$$z = \frac{1}{\varepsilon}\int_f^F \frac{df'}{C(f')}, \tag{B.2c}$$

as always. The coefficients $\sigma(T)$, $b(T)$, $c(T)$, and the constant α are given in terms of $\sigma_0(T)$, $\rho(T)$, $\nu(T)$ and $\tilde{A}(0)$ in [16].

In the SABR and dynamic SABR models, the volatility process $\tilde{A}(T)$ is a log normal Martingale. This may make having extremely large volatilities much too probable[34]. In the mean reverting SABR (λ-SABR) model, the volatility process $\tilde{A}(T)$ is taken to be a mean reverting process to address this issue[8]:

$$d\tilde{F} = \varepsilon\tilde{A}C(\tilde{F})d\tilde{W}_1, \tag{B.3a}$$

$$d\tilde{A} = \varepsilon\lambda(T)[\theta(T) - \tilde{A}]dT + \varepsilon\nu(T)\tilde{A}d\tilde{W}_2, \tag{B.3b}$$

$$d\tilde{W}_1 d\tilde{W}_2 = \rho(T)dT. \tag{B.3c}$$

The analysis in [17] shows that the marginal density $Q(T,F)$ for this model also satisifies the above effective forward equation B.2a-B.2c, at least to within $O\left(\varepsilon^2\right)$.

B.2. Heston-type models. Heston-type models are another popular class of models for managing smile risk[9]. Consider the following volatility model for the forward price $\tilde{F}(T)$ of an asset:

$$d\tilde{F} = \varepsilon\tilde{V}^{1/2}C(\tilde{F})d\tilde{W}_1, \tag{B.4a}$$

$$d\tilde{V} = \varepsilon\kappa(T)\left[\Lambda(T) - \tilde{V}\right]dT + \varepsilon\omega(T)\tilde{V}^{1/2}d\tilde{W}_2, \tag{B.4b}$$

$$d\tilde{W}_1 d\tilde{W}_2 = \rho dT. \tag{B.4c}$$

Unlike the SABR model, the variance per unit time, $\tilde{V}(T)$, diffuses instead of the local volatility, and this variance mean reverts to a equilibrium value $\Lambda(T)$. The above model is a very general Heston model, in that it allows a backbone (local volatility) function $C(F)$, allows the variance $\tilde{V}(T)$ to be correlated with the forward $\tilde{F}(T)$, and all coefficient are allowed to be time-dependent. This model is analyzed in [18], where it is shown that the marginal density satisfies an effective forward equation of the form

$$Q_T = \tfrac{1}{2}\varepsilon^2\theta(T)V\left[\left(1 + 2\varepsilon\frac{b(T)}{\sqrt{V}}z + \varepsilon^2\frac{c(T)}{V}z^2\right)C^2(F)Q\right]_{FF}, \tag{B.5a}$$

$$Q = \delta(F-f) \qquad \text{as } T \to 0, \tag{B.5b}$$

through $O\left(\varepsilon^2\right)$, where V is constant and

$$z = \frac{1}{\varepsilon}\int_f^F \frac{dF'}{C(F')}. \tag{B.5c}$$

Apart from V replacing α^2 and $\theta(T)$ replaceing $\sigma^2(T)$, this is identical to the effective forward equations B.2a-B.2c for the dynamic SABR and λ-SABR models.

B.3. Exponential stochastic volatility models. The exponential stochastic volatility models are[10]

$$d\tilde{F} = \varepsilon\sigma(T,\varepsilon\tilde{Z})e^{\tilde{U}}C(\tilde{F})d\tilde{W}_1, \tag{B.6a}$$

$$d\tilde{U} = \kappa(T)\left[\Lambda(T) - \tilde{U}\right]dT + \varepsilon\nu(T)\,d\tilde{W}_2, \tag{B.6b}$$

$$d\tilde{W}_1 d\tilde{W}_2 = \rho(T)\,dT, \tag{B.6c}$$

where

$$\sigma(T,\varepsilon\tilde{Z}) = 1 + \varepsilon\beta(T)\tilde{Z} + \varepsilon^2\gamma(T)\tilde{Z}^2, \tag{B.6d}$$

with

$$\tilde{Z} = \frac{1}{\varepsilon}\int_f^{\tilde{F}} \frac{dF'}{C(F')}. \tag{B.6e}$$

This model is a hybrid between local and stochastic volatility models, and is often used to combine the local volatility model's ability to calibrate to today's volatility surface with the superior forward volatility surfaces of stochastic vol models. In [19] it is shown that the marginal density $Q(T,F)$ for this model satisfies the effective forward equation B.2a-B.2c through $O\left(\varepsilon^2\right)$.

B.4. Other SABR type models.

B.4.1. Self consistent cross FX models. Let ccy 0 be the base currency (usually USD), and define the forward FX rates to be

$$\tilde{X}_{i0} = \frac{\text{ccy } i}{\text{ccy } 0} = \text{fwd price of 1 unit of ccy } i \text{ in terms of ccy } 0. \tag{B.7}$$

We suppose that the SABR model is used to model the forward FX rates

$$d\tilde{X}_{i0} = \varepsilon\tilde{X}_{i0}\tilde{A}_i dW_i, \tag{B.8a}$$

$$d\tilde{A}_i = \varepsilon\nu_i\tilde{A}_i dZ_i, \tag{B.8b}$$

$$dW_i dZ_i = \rho_{ii} dT, \tag{B.8c}$$

and we also suppose that the model parameters $\alpha_i = \tilde{A}_i(0)$, ρ_{ii}, and ν_i can be obtained by calibrating the SABR smiles to the observed market smiles. The cross-FX rates are then

$$\tilde{X}_{ij} \equiv \frac{\tilde{X}_{i0}}{\tilde{X}_{j0}} = \frac{\text{ccy } i}{\text{ccy } j} = \text{fwd price of 1 unit of ccy } i \text{ in terms of ccy } j. \tag{B.9}$$

Using Itö's lemma shows that

$$\begin{aligned} d\tilde{X}_{ij} &= \frac{d\tilde{X}_{i0}}{\tilde{X}_{j0}} - \frac{\tilde{X}_{i0}}{\tilde{X}_{j0}}\frac{d\tilde{X}_{j0}}{\tilde{X}_{j0}} + \text{ drift terms} \\ &= \varepsilon\tilde{X}_{ij}\left\{\tilde{A}_i dW_i - \tilde{A}_j dW_j\right\} + \text{ drift terms} \end{aligned} \tag{B.10}$$

Since the FX rate $\tilde{X}_{ij}$ is denominated in ccy j, the forward measure for ccy j is the appropriate measure for pricing options on $\tilde{X}_{ij}$. In this measure the forward FX rate $\tilde{X}_{ij}$ is a Martingale, and the drift terms are zero. Thus, the models for all cross FX rates are

(B.11a) $$d\tilde{X}_{ij} = \varepsilon \tilde{X}_{ij} \left\{ \tilde{A}_i dW_i - \tilde{A}_j dW_j \right\},$$

(B.11b) $$d\tilde{A}_i = \varepsilon \nu_i \tilde{A}_i dZ_i,$$

(B.11c) $$d\tilde{A}_j = \varepsilon \nu_j \tilde{A}_j dZ_j,$$

with

(B.11d) $$dW_i dZ_j = \rho_{ij} dT, \qquad dW_i dW_j = C_{ij} dT, \qquad dZ_i dZ_j = R_{ij} dT,$$

for $i, j = 1, 2, \ldots, n$. In [20] it is shown that for any cross rate X, the marginal density $Q(T, X)$ satisfies

(B.12a) $$Q_T = \tfrac{1}{2}\varepsilon^2 \alpha^2 \left[\left(1 + 2\varepsilon bz/\alpha + \varepsilon^2 cz^2/\alpha^2\right) X^2 Q \right]_{XX} \qquad \text{for } T > 0,$$

(B.12b) $$Q = \delta(X - x) \qquad \text{as } T \to 0,$$

through $O\left(\varepsilon^2\right)$, where

(B.12c) $$z \equiv \frac{1}{\varepsilon} \log X/x.$$

This is B.2a-B.2c with $C(X) = X$.

B.4.2. SABR model for baskets and spreads. Let $\tilde{X}(T)$ be the forward value of a basket of n assets,

(B.13) $$\tilde{X} = \sum_{j=1}^{n} \lambda_j \tilde{F}_j,$$

where the forward prices $\tilde{F}(T)$ are governed by *normal* SABR models,

(B.14a) $$d\tilde{F}_j = \varepsilon \tilde{A}_j dW_j,$$

(B.14b) $$d\tilde{A}_j = \varepsilon \nu_j \tilde{A}_j dZ_j,$$

(B.14c) $$dW_i dW_j = c_{ij} dT, \qquad dW_i dZ_j = \rho_{ij} dT, \qquad dZ_i dZ_j = r_{ij} dT,$$

for $i, j = 1, 2, \ldots, n$. This is analyzed in [21], where it is shown that the reduced density $Q(T, X)$ satisfies

(B.15a) $$Q_T = \tfrac{1}{2}\varepsilon^2 \alpha^2 \left[\left(1 + 2\varepsilon bz/\alpha + \varepsilon^2 cz^2/\alpha^2\right) Q \right]_{XX},$$

(B.15b) $$Q(T, X) \to \delta(X - x) \qquad \text{as } \tau \to 0,$$

through $O\left(\varepsilon^2\right)$, with

(B.15c) $$z = \frac{X - x}{\varepsilon}.$$

So, options on the whole basket are governed by the normal SABR model, as long as we can model the individual components via the normal SABR model. It may (or may not) be possible to extend this result to baskets whose components are governed by SABR models with non-zero β's.

A spread $\tilde{X} = \tilde{F}_1 - \tilde{F}_2$ is just a special case of a basket, so options on spreads are also governed by normal SABR models if we can model the individual forwards via normal SABR models.

B.4.3. Generalized ZABR-type models. ZABR-type models,

$$d\tilde{F} = \varepsilon \tilde{A} C(\tilde{F}) d\tilde{W}_1 \tag{B.16a}$$

$$d\tilde{A} = \varepsilon V\left(\tilde{A}\right) d\tilde{W}_2, \tag{B.16b}$$

$$dW_1 dW_2 = \rho\left(\tilde{A}\right) dt, \tag{B.16c}$$

were introduced because in some markets, the vol-of-vol tends to decrease as the volatility $\tilde{A}$ increases, and increase as the volatility decreases. This suggested using a CEV model for the volatility,

$$d\tilde{A} = \varepsilon \nu \tilde{A}^{\gamma} d\tilde{W}_2, \tag{B.17}$$

instead of a log normal model[6]. The analysis in [22] shows that the reduced density $Q(T, F)$ satisfies the effective forward equation B.2a-B.2c through $O\left(\varepsilon^2\right)$.

REFERENCES

[1] P.S. Hagan, D. Kumar, A.S. Lesniewski, and D.E. Woodward, *Managing smile risk*, Wilmott Magazine, September (2002), pp. 84 – 108.

[2] P.S. Hagan, A.S. Lesniewski, and D.E. Woodward, *Probability distribution in the SABR model of stochastic volatility*, working paper (2005)

[3] B Bartlett, *Hedging under SABR Model*, Wilmott Magazine, July (2006), pp. 68 – 70

[4] P.S. Hagan, D. Kumar, A.S. Lesniewski, and D.E. Woodward, *Arbitrage Free SABR*, Wilmott Magazine, January (2013), pp. 1 – 16

[5] P. Balland and Q. Tran, *SABR goes normal,* Risk Magazine, May (2013), pp. 76-81

[6] J. Andreasen and B. N. Huge, *ZABR – Expansions for the Masses,* SSRN: http://ssrn.com/abstract=1980726, Dec (2011)

[7] P. Doust, *No-arbitrage SABR,* J. Comp. Finance, **15** no. 3 (2012), pp. 3 – 31

[8] Pierre Henry-Labordere, *A General Asymptotic Implied Volatility for Stochastic Volatility Models* SSRN: http://ssrn.com/abstract=698601, April (2005)

[9] Steven L.Heston, *A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options*, Rev Fin Studies **6** no 2, (1993), pp. 327–343

[10] L. Scott, *Option Pricing when the Variance changes randomly: Theory, Estimation, and an Application*, J. Fin and Quant Anal **22** (1987), pp. 419-438

[11] P.S. Hagan, D. Kumar, A. Lesniewski, and D.L. Woodward, *Pricing barrier options with the SABR model*, working paper

[12] M.H. Protter and H.F. Weinberger, *Maximum Principles in Differential Equations*, Springer, New York (1984)

[13] B. Dupire, *Pricing with a smile*, Risk, Jan. 1994, pp. 18–20

[14] B. Dupire, *Pricing and hedging with smiles*, in Mathematics of Derivative Securities, M.A.H. Dempster and S. R. Pliska, eds., Cambridge University Press, Cambridge, 1997, pp. 103–111

[15] P.S. Hagan, D. Kumar, A.S. Lesniewski, and D.E. Woodward, *Effective media analysis for stochastic volatility models* , working paper

[16] P.S. Hagan, D. Kumar, A.S. Lesniewski, and D.E. Woodward, *Implied volatility surfaces for the dynamic SABR model*, working paper

[17] P.S. Hagan, D. Kumar, A.S. Lesniewski, and D.E. Woodward, *Effective forward equations for the mean reverting SABR model*, working paper

[18] P.S. Hagan, D. Kumar, A.S. Lesniewski, and D.E. Woodward, *Effective forward equations for the generalized Heston model* , working paper

[19] P.S. Hagan, D. Kumar, A.S. Lesniewski, and D.E. Woodward, *Effective forward equations for exponential stochastic volatility model* , working paper

[20] P.S. Hagan, D. Kumar, A.S. Lesniewski, and D.E. Woodward, *Self consistent cross FX smile models*, working paper

[21] P.S. Hagan, D. Kumar, A.S. Lesniewski, and D.E. Woodward, *Self consistent smile models for baskets and spreads*, working paper

[22] P.S. Hagan, D. Kumar, A.S. Lesniewski, and D.E. Woodward, *Implied volatility smiles for ZABR-type models*, working paper

[23] John C. Neu, *thesis*, California Institute of Technology, (1978)

[24] J.M. Harrison and S. Pliska, *Martingales and stochastic integrals in the theory of continuous trading*, Stoch. Proc. Appl, 11 (1981), pp. 215-260.
[25] J.M. Harrison and D. Krebs, *Martingales and arbitrage in multiperiod securities markets*, J. Econ. Theory, 20 (1979), pp. 381-408.
[26] I. Karatzas, J.P. Lehoczky, S.E. Shreve, and G.L. Xus, *Martingale and duality methods for utility maximization in an incomplete market*, SIAM J. Control Optim, 29 (1991), pp. 702–730.
[27] J. Michael Steele, *Stochastic Calculus and Financial Applications*, Springer, 2001
[28] I. Karatzas and S. Shreve, *Brownian Motion and Stochastic Calculus*, Springer, 1988.
[29] B. Okdendal, *Stochastic Differential Equations*, Springer, 1998.
[30] M. Musiela and M. Rutkowski, *Martingale Methods in Financial Modelling*, Springer, 1998.
[31] Patrick S. Hagan, *Convexity conundrums: Pricing CMS swaps, caps, and floors*, Wilmott, March 2003, pp. 38–44
[32] Patrick S. Hagan, *Conservative schemes for solving one spatial dimension PDEs*, working paper
[33] P. Wilmott, S. Howison, J. DeWynne, *Option Pricing: Mathematical Models and Computation*, Oxford Financial Press, 1993
[34] Jim Gatheral, *The Volatility Surface*, Wiley, 2006
[35] P.S. Hagan, D. Kumar, A.S. Lesniewski, and D.E. Woodward, *Cutailing the high strike volatility and taming the CMS deals*, working paper (2015).