Implied Remaining Variance in Derivative Pricing

PETER CARR AND JIAN SUN

PETER CARR
is the executive director of the Masters in Math Finance Program at New York University in New York, NY.
pcarr@nyc.rr.com

JIAN SUN
is an executive director at a large financial institution in New York, NY.
jiansun88@gmail.com

Consider a call swaption maturing at time $T > 0$ written on an underlying swap maturing at a later time $T' > T$. Let $A_t(T,T')$ be the spot price at time $t \in [0,T]$ of a forward-starting annuity whose payments begin at T and end at T'. Let $F_t(T,T')$ be the forward swap rate at time $t \in [0,T]$. In what follows, we fix both T and T', so we henceforth denote the forward-starting annuity value by just A_t and likewise denote the forward swap rate by just F_t. The call swaption's time T payoff in units of the forward-starting annuity is $(F_T-K)^+$, where $K > 0$ is the strike price. It is well known that the absence of arbitrage between swaptions of maturity T and swaps maturing at T and T' implies the existence of an equivalent martingale measure Q_S, commonly called forward swap measure (see, e.g., Wu [2009]). Under the forward swap measure Q_S, the forward swap rate F is a martingale. One of the simplest possible specifications of this martingale is driftless GBM, i.e.,

$$dF_t = \sigma F_t dW_t, \quad t \in [0,T] \tag{1}$$

where σ is a real constant called volatility and W is standard Brownian motion under Q_S. This model is commonly referred to as Black's model.

With the parameter σ numerically determined at time 0, we can define a non-negative deterministic process $R_t \equiv \sigma^2(T-t)$, $t \in [0,T]$ called remaining variance. This process starts at $\sigma^2 T$ at time $t = 0$ and then declines linearly towards zero, reaching zero at $t = T$. Let C_t be the theoretical value at time $t \in [0,T]$ of a call swaption, measured in forward-starting annuities. In the above Black model, the non-negative process C_t depends only on the contemporaneous forward swap rate F_t, the strike price K, and the remaining variance R_t. Let $C(F,K,R)$: $(R^+)^3 \mapsto R^+$ be the Black call formula, which relates the call's theoretical value to these three variables:

$$C(F,K,R) \equiv FN(d_1(k,R)) - KN(d_2(k,R)) \tag{2}$$

where:

$$d_2(k(F,K),R) \equiv \frac{-k}{\sqrt{R}} - \frac{\sqrt{R}}{2},$$
$$d_1(k(F,K),R) \equiv d_2(k(F,K),R) + \sqrt{R},$$
$$k(F,K) \equiv \ln\left(\frac{K}{F}\right) \tag{3}$$

Then in the Black model, the arbitrage-free call value process is:

$$C_t = C(F_t,K,R_t), \qquad t \in [0,T] \tag{4}$$

In an arbitrage-free market consisting of a single swaption maturing at T, the parameter

σ can always be chosen to achieve a perfect fit to that swaption. When σ is determined in this way, as opposed to some other method such as using time series of the underlying, we call it implied vol and denote it σ^{BS}. In a market with two or more swaptions all maturing at T, one can numerically determine the value of σ that minimizes pricing error. Unfortunately, this minimum pricing error is widely considered to be intolerably large for market makers and intolerably persistent for traders.

This has led to the development of more general stochastic volatility (SV) models for which the forward swap rate F obeys:

$$dF_t = \sigma_t F_t dW_t, \quad t \in [0,T] \tag{5}$$

where the instantaneous volatility σ_t is a univariate time homogeneous diffusion:

$$d\sigma_t = \alpha(\sigma_t)dt + \beta(\sigma_t)dZ_t, \quad t \in [0,T] \tag{6}$$

Here, $\alpha(\sigma)$ and $\beta(\sigma)$ are both functions mapping R to R, while Z is a standard Brownian motion under Q_S. The increments in Z have constant correlation with those of W, i.e,:

$$dW_t dZ_t = \rho dt, \quad t \in [0,T] \tag{7}$$

where the constant $\rho \in [-1,1]$. Examples of this type of SV model include Heston [1993] and Hull and White [1987].

As long as one can dynamically trade in both a forward swap and a swaption, the price of every other co-terminal swaption is determined by the stochastic volatility model. Moreover, for fixed T and T', the arbitrage-free price process for a call swaption strike profile $C_t^{sv}(K), K > 0$ is determined via:

$$C_t^{sv}(K) = C^{sv}(F_t, K, \sigma_t, T-t), \quad t \in [0,T] \tag{8}$$

for some function C^{sv} of its four indicated arguments and parameters of the σ process (including those governing correlation).

Once a call price $C_t^{sv}(K)$ has been determined, one can also determine the corresponding Black implied volatility $I_t(K)$. This implied volatility arises by equating the Black model call value to the SV model call value and numerically solving for the σ parameter in the former model value. Once the Black implied volatility curve $I_t(K)$ has been determined, one can also determine a concept we call the "remaining implied variance strike profile, denoted by $\omega_t(K)$, $K > 0$, and defined by:

$$\omega_t(K) = I_t^2(K)(T-t), \quad K > 0, t \in [0,T] \tag{9}$$

Alternatively, one can go directly from the arbitrage-free curve of call prices $C_t^{sv}(K), K > 0$ to the contemporaneous and co-terminal remaining implied variance strike profile via the following implicit definition:

$$C_t^{sv}(K) = C(F_t, K, \omega_t(K)) \tag{10}$$

$$\equiv F_t N(d_1(F_t, K, \omega_t(K)) - KN(d_2(F_t, K, \omega_t(K)) \tag{11}$$

for $t \geq 0$, $K > 0$. The difference between the R_t process used in Equation (4) and the process $\omega_t(K)$ just defined is that R_t is a scalar process independent of strike K, while $\omega_t(K)$ is a stochastic curve, indexed by strike $K > 0$. This difference arises because R_t is defined as remaining implied variance at time t when the Black model holds, while $\omega_t(K)$ is defined as remaining implied variance at time t when an SV model holds.

In general, $\omega_t(K)$ depends not only on the indicated variables t and K, but also on F_T, σ_t and T–t. The selection of a drift coefficient $\alpha(\sigma)$ and a diffusion coefficient $\beta(\sigma)$ in Equation (6) determines an SV model. This selection determines not only the initial arbitrage-free price of all call swaptions $C_0^{sv}(K), K > 0$ but also the dynamics of call values:

$$dC_t^{sv}(K) = \frac{\partial C^{sv}}{\partial F}(F_t, K, \sigma_t, T-t)\sigma_t F_t dW_t + \frac{\partial C^{sv}}{\partial \sigma}(F_t, K, \sigma_t, T-t)\beta(\sigma_t)dZ_t \tag{12}$$

for $t \in [0,T]$. It follows from Equation (11) that the selection of $\alpha(\sigma)$ and $\beta(\sigma)$ also determines the initial remaining implied variance strike profile $\omega_0(K)$, $K > 0$ as well as its dynamics under Q_S.

In this article, we try to bypass the selection of $\alpha(\sigma)$ and $\beta(\sigma)$ by directly modeling the Q_S dynamics of the remaining implied variance strike profile. We recognize that such an approach is fraught with the danger that we may inadvertently introduce cross strike arbitrage. We proved later in this article that we can select parameters to avoid arbitrage. The main advantage of our approach versus a classical SV model is that we will be able to produce an explicit formula relating initial remaining implied variance $\omega_0(K)$ to strike price K. Inverting Equation (9) for Black implied vol:

$$I_t(K) = \sqrt{\frac{\omega_t(K)}{T-t}}, \quad K > 0, t \in [0,T] \tag{13}$$

and setting $t = 0$ means that we also have an explicit formula relating initial implied vol $I_0(K)$ to strike price K. For SV models, there are no explicit exact formulas for $I_0(K)$, but there are explicit approximations. Unfortunately, these explicit approximations are relatively complicated and they produce arbitrage at extreme strikes.

Hence, the main idea of this article is to replace the Q_S dynamics of instantaneous volatility in Equation (6) with the following Q_S dynamics for the curve $\omega_t(K)$:

$$d\omega_t(K) = a(\omega_t(K))\sigma_t^2 dt + b(\omega_t(K))\sigma_t dZ_t, \quad K > 0, t \in [0,T] \tag{14}$$

where Z is a Q_S standard Brownian motion. As the notation in Equation (14) indicates, $a(\omega)$ and $b(\omega)$ are two real-valued functions which depend *only* on $\omega \geq 0$, i.e., $a(\omega)$: $R^+ \mapsto R$ and $b(\omega)$: $R^+ \mapsto R$. As a result, at each fixed $t \in [0,T]$, all functional dependence of the random variables $a(\omega_t(K))$ and $b(\omega_t(K))$ on the contemporaneous or past levels of the processes σ, F, Z, or W can only occur through $\omega_t(K)$. The reason for the very specific way that the process σ_t enters into the dynamics implies that at each $K > 0$, the scalar stochastic process $\omega_t(K)$ has the same laws as a univariate time-homogeneous diffusion running on a stochastic clock given by the quadratic variation of $\ln F$. The virtue of limiting σ's role to determining quadratic variation will soon become apparent. Note that Z is a single standard Brownian motion that drives the entire remaining implied variance curve.

The correlation between increments dW_t and dZ_t is again given by ρdt, where ρ is a constant in $[-1,1]$. We restrict the functions $a(\omega)$ and $b(\omega)$ appearing in Equation (14), so that if we were given the initial profile $\omega_0(K)$, $K > 0$, then the solution to the SDE is unique. Carr and Sun [2007] use a specification similar to Equation (14) for the remaining variance in a variance swap. See Carr and Wu [2012] for an alternative specification of implied vol dynamics.

In our problem, we are not actually given the initial profile $\omega_0(K)$, $K > 0$. Rather, the goal will be to determine the initial profile when we are only given a few market-implied volatilities. Once one specifies some parametric form for the functions $a(\omega)$ and $b(\omega)$, we will show that a resulting parametric form for the initial profile is determined. Once the parameters in the parametric form for $a(\omega)$ and $b(\omega)$ are numerically determined, the initial profile is determined as well. In other words, the dynamics of each remaining implied variance over time will be shown to determine how the initial remaining implied variance depends on strike price. A well-known example is the Black model whose dynamics are $b = 0$, $a = 1$, and $\sigma_t = 1$. In this case, the initial remaining implied variance has no dependence on strike price since $\omega_0(K) = T$.

Under forward swap measure Q_S, the call price process $C_t^{sv}(K)$ is a martingale. However, we recall that by the definition of ω_t:

$$C_t^{sv}(K) = C(F_t, K, \omega_t(K)), \quad t \in [0,T] \tag{15}$$

When we use Itô's formula on Equation (15) in conjunction with our dynamical assumption Equation (14), the zero-drift condition implies that at each time $t \in [0,T]$, the curve $\omega_t(K)$ must satisfy:

$$\begin{aligned}
-a(\omega_t(K))\sigma_t^2 \frac{\partial C}{\partial \omega}(F_t, K, \omega_t(K))
&= \frac{\sigma_t^2}{2} F_t^2 \frac{\partial^2 C}{\partial F^2}(F_t, K, \omega_t(K)) \\
&\quad + \sigma_t^2 \rho b(\omega(K)) F_t \frac{\partial^2 C}{\partial F \partial \omega}(F_t, K, \omega_t(K)) \\
&\quad + \frac{\sigma_t^2}{2} b^2(\omega(K)) \frac{\partial^2 C}{\partial \omega^2}(F_t, K, \omega_t(K))
\end{aligned} \tag{16}$$

Note that setting σ_t to a deterministic function of time and setting $b(\omega) = 0$ in Equation (14) causes $\omega_t(K)$ to evolve deterministically, as it would in a time-dependent Black model. Further setting σ_t to a constant and also setting the function $a(\omega) = -1$ causes $\omega_t(K)$ to decline linearly toward zero, as it would in a constant-parameter Black model. Now suppose that the dynamics of $\omega(K)$ are governed by Equation (6) with $b(\omega_t(K)) \neq 0$. Then it is still the case that the function $C(F,K,R)$ defined by Equation (2) solves the PDE:

$$\frac{\partial C}{\partial \omega}(F,K,R) = \frac{F^2}{2}\frac{\partial^2 C}{\partial F^2}(F,K,R) \quad F>0, K>0, R>0 \tag{17}$$

Were the Black model holding, then the absence of arbitrage forces the rate at which convexity gains $\frac{F^2}{2}\frac{\partial^2 C}{\partial F^2}(F,K,R)$ increase to be offset exactly by positive exposure $\frac{\partial C}{\partial \omega}(F,K,R)$ to the declining process $\omega_t(K) = \sigma^2(T-t)$. The partial derivative $\frac{\partial C}{\partial R}(F,K,R)$ is worthy of a greek letter, so we christen it iota to remind us that it is the sensitivity to implied remaining variance. Were the Black model holding, then Equation (17) implies that the rate $\frac{F^2}{2}\frac{\partial^2 C}{\partial F^2}(F,K,R)$ at which convexity profits increase is offset exactly by positive exposure $\frac{\partial C}{\partial \omega}(F,K,R)$ to the declining process $\omega_t = \sigma^2(T-t)$. This is a necessary consequence of the absence of arbitrage whenever the Black model is holding and hence $b(\omega_t(K)) = 0$.

However, when $b(\omega_t(K)) \neq 0$, then Equation (16) implies that the stochastic process $\frac{\partial C}{\partial \omega}(F_t,K,\omega_t(K))$, $t \in [0,T]$ does not equal the stochastic process $\frac{F_t^2}{2}\frac{\partial^2 C}{\partial F^2}(F_t,K,\omega_t(K)), t \in [0,T]$. Rather, to avoid arbitrage, the profits arising from dollar gamma and the other two second-order greeks are offset exactly by the process $a(\omega_t(K))\sigma_t^2 \frac{\partial C}{\partial \omega}(F_t,K,\omega_t(K)), t \in [0,T]$.

Solving Equation (16) for the drift process $a(\omega_t(K))$:

$$a(\omega_t(K)) = -\left[\begin{array}{c} \frac{F_t^2}{2}\frac{\partial^2 C}{\partial F^2}\Big/\frac{\partial C}{\partial \omega} + \rho b(\omega_t(K))F_t \\ \frac{\partial^2 C}{\partial F \partial \omega}\Big/\frac{\partial C}{\partial \omega} + \frac{b^2(\omega_t(K))}{2}\frac{\partial^2 C}{\partial \omega^2}\Big/\frac{\partial C}{\partial \omega} \end{array}\right] \tag{18}$$

where the arguments $(F_t, K, \omega_t(K))$ of C have been suppressed. From Equation (18), we see that no arbitrage forces the drift in remaining implied variance under Q_S to be fully explained by three greek ratios rather than one. Notice that if $b(\omega_t(K) = 0$, then the drift $a(\omega_t(K))$ is negatively proportional to the ratio of dollar gamma to $\frac{F_t^2}{2}\frac{\partial^2 C}{\partial F^2}$ to $\frac{\partial C}{\partial \omega}$. This sole dependence arises in the Black model, but one may be surprised to learn that deterministically time changed GBM is actually not necessary for this result to hold at the single maturity date T. As long as the stochastic instantaneous volatility process is such that $<\ln F>_T$ is known at every time $t \in [0,T]$, the remaining implied volatility $\omega_t(K)$ has no local martingale component ($b(\omega) = 0$), so the drift $a(\omega_t(K))$ is negatively proportional to the ratio of dollar gamma to $\frac{F_t^2}{2}\frac{\partial^2 C}{\partial F^2}$ to $\frac{\partial C}{\partial \omega}$. This ratio is actually constant, so ω_t does not depend on K.

A standard Heath Jarrow Morton (Heath et al. [1991]) (HJM) style approach is to specify both the initial level of a curve and specify the second-order structure of a curve and the two specifications jointly determine the drift of the curve. Here, we will instead specify both the first- and second-order structure of a curve and and then determine the initial level of the curve. To ensure the existence and uniqueness of a solution, we will proceed parametrically. Hence, given just a few market-implied volatilities, one can choose some lower order parametrization of the functions $a(\omega)$ and $b(\omega)$, minimize squared error to determine the parameters, and then compute the initial implied remaining variance curve. As a consequence of the particular way that the process σ_t enters the SDE Equation (14), the choice of parametrization for the functions $a(\omega)$ and $b(\omega)$ will have no impact on the instantaneous volatility process σ_t.

GREEKS COMPUTATION

Consider the Black call swaption pricing formula in Equation (2) with R replaced by ω and where for notational simplicity, we suppress the dependence of the function $k(F,K) \equiv \ln(K/F)$ on F and K:

$$C(F,K,\omega) \equiv FN(d_1(k,\omega)) - KN(d_2(k,\omega)) \tag{19}$$

where:

$$d_2(k,\omega) \equiv \frac{-k}{\sqrt{\omega}} - \frac{\sqrt{\omega}}{2}, \quad d_1(k,\omega) \equiv d_2(k,\omega) + \sqrt{\omega} \tag{20}$$

We now use it to compute the four Greeks in Equation (18).

$$\frac{\partial}{\partial \omega}C(F,K,\omega) = \frac{KN'(d_2(k,\omega))}{2\sqrt{\omega}} \tag{21}$$

$$\frac{F^2}{2}\frac{\partial^2}{\partial F^2}C(F,K,\omega) = \frac{KN'(d_2(k,\omega))}{2\sqrt{\omega}} \tag{22}$$

$$F\frac{\partial^2}{\partial\omega\partial F}C(F,K,\omega)=\frac{KN'(d_2(k,\omega))}{2\sqrt{\omega}}\left[-\frac{d_2(k,\omega)}{\sqrt{\omega}}\right] \quad (23)$$

$$\frac{1}{2}\frac{\partial^2}{\partial\omega^2}C(F,K,\omega)=\frac{KN'(d_2(k,\omega))}{2\sqrt{\omega}}\left[\frac{d_1(k,\omega)d_2(k,\omega)-1}{4\omega}\right] \quad (24)$$

Therefore:

$$\frac{F^2}{2}\frac{\partial^2 C}{\partial F^2}\Big/\frac{\partial C}{\partial\omega}=1 \quad (25)$$

$$F\frac{\partial^2 C}{\partial\omega\partial F}\Big/\frac{\partial C}{\partial\omega}=-\frac{d_2(k,\omega)}{\sqrt{\omega}} \quad (26)$$

$$\frac{1}{2}\frac{\partial^2 C}{\partial\omega^2}\Big/\frac{\partial C}{\partial\omega}=\frac{d_1(k,\omega)d_2(k,\omega)-1}{4\omega} \quad (27)$$

Substituting in the definitions in Equation (20) of $d_1(k,\omega)$ and $d_2(k,\omega)$ in Equations (25) to (27):

$$\frac{F^2}{2}\frac{\partial^2 C}{\partial F^2}\Big/\frac{\partial C}{\partial\omega}=1 \quad (28)$$

$$F\frac{\partial^2 C}{\partial\omega\partial F}\Big/\frac{\partial C}{\partial\omega}=\frac{k}{\omega}+\frac{1}{2} \quad (29)$$

$$\frac{1}{2}\frac{\partial^2 C}{\partial\omega^2}\Big/\frac{\partial C}{\partial\omega}=\frac{\frac{k^2}{\omega}-\frac{\omega}{4}-1}{4\omega} \quad (30)$$

Substituting Equation (28) to (30) into Equation (18), we end up with the following simple form for the determinants of $a(\omega)$:

$$a(\omega)=-\left\{1+\frac{\rho b(\omega)}{2}-\left(\frac{1}{\omega}+\frac{1}{4}\right)\frac{b^2(\omega)}{4}+\frac{\rho b(\omega)}{\omega}k+\frac{b^2(\omega)}{4\omega^2}k^2\right\} \quad (31)$$

The next step is to pick a parametric functional form for $a(\omega)$ and $b(\omega)$, which ideally leads to a simple and yet realistic closed form solution for $\omega(K)$.

A PARAMETRIC FUNCTIONAL FORM FOR a(ω) AND b(ω)

We now choose a simple functional form for the functions $a(\omega)$ and $b(\omega)$ in Equation (31). We assume the simple forms:

$$a(\omega)=-a_1\omega+a_0-1 \quad (32)$$

$$b(\omega)=b\,\omega \quad (33)$$

where $a_1\in R$, while $a_1>0$ and $b>0$. Substituting Equations (32) and (33) in Equation (31) leads to the following quadratic equation for ω:

$$\frac{b^2}{16}\omega^2+\left(\frac{b^2}{4}-\frac{\rho b}{2}+a_1\right)\omega-\left[a_0+\rho b\ln\left(\frac{K}{F}\right)+\frac{b^2}{4}\ln^2\left(\frac{K}{F}\right)\right]=0 \quad (34)$$

When:

$$\frac{b^2}{4}-\frac{\rho b}{2}+a_1>0 \quad (35)$$

and

$$a_0>\rho^2 \quad (36)$$

then the quadratic equation (Equation (34)) always has a positive solution. This is because the linear coefficient of the quadratic equation is positive, but the zero order term of the quadratic equation

$$a_0+\rho b\ln\left(\frac{K}{F}\right)+\frac{b^2}{4}\ln^2\left(\frac{K}{F}\right)>0$$

universally which is straightforward by checking its discriminant as a quadratic function of $\ln(\frac{K}{F})$.

If we divide Equation (34) by b^2, we arrive at a quadratic equation for ω determined by just three parameters, namely ρ, $\frac{a_1}{b_1}$, and $\frac{a_0}{b}$. Alternatively, we can just set $b=1$ and vary a_1,a_0,ρ to meet the market. As a result, the dynamical specification of F and $\omega(K)$ over the time interval $t\in[0,T]$ is given by:

$$dF_t=\sigma_t F_t dW_t \quad (37)$$

$$d\omega_t(K)=-(1-a_0+a_1\omega_t(K))\sigma_t^2dt+\omega_t(K)\sigma_t dZ_t \quad (38)$$

$$dW_t dZ_t=\rho dt \quad (39)$$

Using the quadratic root formula, the positive root of (34) with $b=1$ is:

$$\omega(K) = \sqrt{2}\rho - 2 - a_1 + \sqrt{D(K/F)} > 0 \qquad (40)$$

where the discriminant $D(K/F)$ is given by:

$$\left(\sqrt{2}\rho - 2 - a_1\right)^2 + 2a_0 + 4\sqrt{2}\rho\ln(K/F) + 4\ln^2(K/F) \qquad (41)$$

Note that the asymptotics have implied variance ω linear in log moneyness $\ln(K/F)$. According to the Lee [2004] moment formula, these asymptotics are arbitrage-free. As we vary the elements of the triple (a_1, a_0, ρ), we are able to generate various shapes of the implied volatility curve. In general a_1,a_0 jointly control the curvature and height, while ρ controls the slope of the curve. We find that this three-parameter model provides a good fit to co-terminal market swaption-implied volatilities.

THE LINK WITH GVV

We now explore the connection of this model with GVV model (Arslan et al. [2009]). We keep the notation that

$$\omega = I^2(T-t)$$

and we have the following set of equations

$$\frac{\partial C}{\partial \omega} = -\frac{1}{I^2}\frac{\partial C}{\partial t}$$

$$\frac{\partial^2}{\partial \omega^2} = \left(\frac{\partial^2 C}{\partial I^2} + 2(t-t)\frac{1}{I^2}\frac{\partial C}{\partial t}\right)\frac{1}{4I^2(t-t)^2}$$

$$\frac{\partial^2 C}{\partial F\,\partial \omega} = \frac{1}{2I(T-t)}\frac{\partial^2 C}{\partial F\,\partial I}$$

with these equations and plug into Equation (16) and using the parametrization in Equation (32), we have the following reorganized equation:

$$\frac{1}{2}F^2\frac{\partial^2 C}{\partial F^2} + \frac{1}{2}\rho FBI\frac{\partial^2 C}{\partial F\,\partial I} + \frac{1}{8}b^2I^2\frac{\partial^2 C}{\partial I^2} + \frac{1}{4}b^2(T-t)\frac{\partial C}{\partial t} - a_1(T-t)$$

$$\frac{1}{2}F^2\frac{\partial^2 C}{\partial F^2} + \frac{1}{2}\rho FbI\frac{\partial^2 C}{\partial F\,\partial I} + \frac{1}{8}b^2I^2\frac{\partial^2 C}{\partial I^2} = \left(a_1(T-t) - \frac{1}{4}b^2(T-t)\right)\frac{\partial C}{\partial t}$$

We interpret this as following: the option theta is a linear combination of dollar vega, dollar vanna, and dollar volga. This is in line with the GVV model.

NO ARBITRAGE

In this section, we investigate the parameters we have chosen in Equation (32), which will guarantee no arbitrage. First of all let us restate the parametrization.

$$a(\omega) = -a_1\omega + a_0 - 1 \qquad (42)$$

$$b(\omega) = b\,\omega \qquad (43)$$

with restrictions

$$\frac{b^2}{4} - \frac{\rho b}{2} + a_1 > 0 \qquad (44)$$

and

$$a_0 > \rho^2 \qquad (45)$$

We have shown that under these constraints, we should always have positive solutions for $\omega(K)$. In the following analysis, we always let $b = 1$ and as a consequence we have the constraint

$$\frac{1}{4} - \frac{\rho}{2} + a_1 > 0$$

As a function of strike K, we define

$$B(K) = C(K, \omega(K)) \qquad (46)$$

where C is the Black-Scholes call price with strike at K and implied remaining variance at ω. Its value is given by

$$C(K,\omega) = FN(d_1) - KN(d_2)$$

and d_1, d_2 given by

$$d_1 = \frac{\ln(F/K) + \frac{1}{2}\omega}{\sqrt{\omega}}, \quad d_2 = \frac{\ln(F/K) - \frac{1}{2}\omega}{\sqrt{\omega}}$$

To show there is no arbitrage between the calls, we must prove the monotonicity inequality

$$\frac{\partial B}{\partial K} < 0$$

and convexity inequality

$$\frac{\partial^2 B}{\partial K^2} > 0$$

for every $K > 0$. For this purpose, we calculate

$$\frac{\partial B}{\partial K} = \frac{\partial C}{\partial \omega}\omega_K$$

and

$$\frac{\partial^2 B}{\partial K^2} = \frac{\partial^2 C}{\partial K^2} + 2\frac{\partial^2 C}{\partial K \partial \omega}\omega_K + \frac{\partial^2 C}{\partial \omega^2}\omega_K^2 + \frac{\partial C}{\partial \omega}\omega_{KK}$$

We can calculate each term

$$\frac{\partial C}{\partial \omega} = \frac{K}{2\sqrt{\omega}}N'(d_2)$$
$$\frac{\partial^2 C}{\partial K^2} = \frac{1}{K\sqrt{\omega}}N'(d_2)$$
$$\frac{\partial^2 C}{\partial K \partial \omega} = \frac{1}{2\sqrt{\omega}}N'(d_2) + \frac{d_2}{2\omega}N'(d_2)$$
$$\frac{\partial^2 C}{\partial \omega^2} = \frac{K}{\sqrt{\omega}}N'(d_2)\left[\frac{d_1 d_2 - 1}{4\omega}\right]$$

therefore

$$\frac{\partial^2 C}{\partial K^2}\Big/\frac{\partial C}{\partial \omega} = \frac{2}{K^2}$$
$$\frac{\partial^2 C}{\partial K \partial \omega}\Big/\frac{\partial C}{\partial \omega} = \frac{1}{K} + \frac{d_2}{\sqrt{\omega}K}$$
$$\frac{\partial^2 C}{\partial \omega^2}\Big/\frac{\partial C}{\partial \omega} = \left[\frac{d_1 d_2 - 1}{2\omega}\right]$$

The monotonicity inequality is equivalent to

$$\frac{1}{2\sqrt{\omega}}N'(d_2)\omega_K \leq N(d_2) \qquad (47)$$

The convexity inequality is equivalent to

$$\frac{2}{K^2} + 2\left(\frac{1}{K} + \frac{d_2}{\sqrt{\omega}K}\right)\omega_K + \frac{d_1 d_2 - 1}{2\omega}\omega_K^2 + \omega_{KK} > 0 \qquad (48)$$

We know that ω satisfies the equation

$$\frac{1}{16}\omega^2 + \left(\frac{1}{4} - \frac{\rho}{2} + a_1\right)\omega - \left(\frac{1}{4}\ln^2(F/K) + \rho\ln(F/K) + a_0\right) = 0$$

We denote

$$\beta = \frac{1}{4} - \frac{\rho}{2} + a_1, \quad \alpha = \frac{1}{4}\ln^2(F/K) + \rho\ln(F/K) + a_0$$

and by the quadratic solution formula

$$\omega = 8\left(-\beta + \sqrt{\beta^2 + \frac{1}{4}\alpha}\right)$$

We can also calculate its first and second derivatives

$$\omega_K = -\frac{1}{K}\frac{\rho + \frac{1}{2}\ln(F/K)}{\sqrt{\beta^2 + \frac{1}{4}\alpha}}$$
$$\omega_{KK} = \frac{1}{K^2}\frac{\rho + \frac{1}{2}\ln(F/K) + \frac{1}{2}}{\sqrt{\beta^2 + \frac{1}{4}\alpha}} - \frac{1}{8}\frac{1}{K^2}\frac{(\rho + \frac{1}{2}\ln(F/K))^2}{(\beta^2 + \frac{1}{4}\alpha)^{3/2}}$$

Plugging into Equation (47), we have the inequality

$$-\frac{1}{2\sqrt{\omega}}\frac{\rho + \frac{1}{2}\ln(F/K)}{\sqrt{\beta^2 + \frac{1}{4}\alpha}}N'(d_2) \leq N(d_2) \qquad (49)$$

Plugging into Equation (48), we have the following inequality, which is equivalent to the convexity requirement

$$2 + \frac{1}{2}\frac{1}{\sqrt{\beta^2 + \frac{1}{4}\alpha}} + \frac{d_1 d_2 - 1}{2\omega}\frac{(\rho + \frac{1}{2}\ln(F/K))^2}{\beta^2 + \frac{1}{4}\alpha}$$
$$\geq 2\frac{\ln(F/K)}{\omega}\frac{\rho + \frac{1}{2}\ln(F/K)}{\sqrt{\beta^2 + \frac{1}{4}\alpha}} + \frac{1}{8}\frac{(\rho + \frac{1}{2}\ln(F/K))^2}{(\beta^2 + \frac{1}{4}\alpha)^{3/2}} \qquad (50)$$

To simplify, if we let $x = \ln(F/K)$, we can shorten these two inequalities to

$$-\frac{1}{2\sqrt{\omega}}\frac{\rho + \frac{1}{2}x}{\sqrt{\beta^2 + \frac{1}{4}\alpha}}N'(d_2) \leq N(d_2)$$

and

$$2+\frac{1}{2}\frac{1}{\sqrt{\beta^2+\frac{1}{4}\alpha}}+\frac{d_1d_2-1}{2\omega}\frac{(\rho+\frac{1}{2}x)^2}{\beta^2+\frac{1}{4}\alpha}$$

$$\geq 2\frac{x}{\omega}\frac{\rho+\frac{1}{2}x}{\sqrt{\beta^2+\frac{1}{4}\alpha}}+\frac{1}{8}\frac{(\rho+\frac{1}{2}x)^2}{(\beta^2+\frac{1}{4}\alpha)^{3/2}}$$

Lemma 1. We have the following limit

$$\lim_{K\to 0}\frac{\partial B}{\partial K}=-1$$

$$\lim_{K\to +\infty}\frac{\partial B}{\partial K}=0$$

i.e., the first derivative is -1 as K goes to zero and approaches zero as K goes to infinity.

Proof. Again let us use $x = \ln(F/K)$. Because $\alpha=\frac{1}{4}x^2+\rho x+a_0$ and

$$\omega=8\left(-\beta+\sqrt{\beta^2+\frac{1}{4}\alpha}\right)=\frac{2\alpha}{-\beta+\sqrt{\beta^2+\frac{1}{4}\alpha}}$$

we see that $\omega\sim 2\,|\,x\,|$ as $x\to\infty$. This proves that

$$\lim_{x\to+\infty} d_2=+\infty,\quad \lim_{x\to-\infty} d_2=-\infty$$

It is straightforward to check that as $K\to+\infty$, $x\to-\infty$ we have

$$d_2\to+\infty,\quad N(d_2)\to 0$$

and as $K\to+\infty$, $x\to-\infty$ we have

$$d_2\to-\infty,\quad N(d_2)\to 1$$

In the inequality (Equation (48)), we have three terms on the left and two on the right. We want to show that one term on the left always dominates one term on the right.

Lemma 2. The inequality

$$\frac{1}{2}\frac{1}{\sqrt{\beta^2+\frac{1}{4}\alpha}}\geq\frac{1}{8}\frac{(\rho+\frac{1}{2}x)^2}{(\beta^2+\frac{1}{4}\alpha)^{3/2}}$$

is true everywhere.

Proof. After cancellation, we just need to show

$$4\left(\beta^2+\frac{1}{4}\alpha\right)\geq\left(\rho+\frac{1}{2}x\right)^2$$

But we know that

$$\alpha=\frac{1}{4}x^2+\rho x+a_0=\left(\rho+\frac{1}{2}x\right)^2+a_0-\rho^2\geq\left(\rho+\frac{1}{2}x\right)^2$$

therefore, the stated inequality should be true everywhere.

Because of the lemma above, we are left to show that

$$2+\frac{d_1d_2-1}{2\omega}\frac{(\rho+1/2x)^2}{\beta^2+1/4\alpha}\geq 2\frac{x}{\omega}\frac{\rho+\frac{1}{2}x}{\sqrt{\beta^2+1/4\alpha}} \tag{51}$$

by working out d_1d_2–1, we have the following equivalent inequality

$$2+\frac{x^2-\frac{1}{4}\omega^2-\omega}{2\omega^2}\frac{(\rho+1/2x)^2}{\beta^2+1/4\alpha}\geq 2\frac{x}{\omega}\frac{\rho+\frac{1}{2}x}{\sqrt{\beta^2+1/4\alpha}} \tag{52}$$

but we will see further constraints on a_1, a_0, ρ will be needed for this to be true. We now state the local convexity result.

Theorem 3. If the parameters a_1,a_0,ρ satisfies that

$$32Y^2(Y-\beta)>(1+2(Y-\beta))\rho^2$$

where

$$Y=\sqrt{\beta^2+\frac{1}{4}a_0}$$

the inequality (Equation (50)) will be true around $K = F$.

Proof. At $K = F$, we have $x = 0$, therefore by inequality Equation (52) we are left to show

$$2+\frac{d_1d_2-1}{2\omega}\frac{(\rho+\frac{1}{2}x)^2}{\beta^2+\frac{1}{4}\alpha}>0$$

If d_1d_2–1 > 0 obviously this is true and if otherwise $d_1d_2 < 1$ this is equivalent to

$$2 \geq \frac{1-d_1d_2}{2\omega}\frac{(\rho+\frac{1}{2}x)^2}{\beta^2+\frac{1}{4}\alpha}$$

and when $x = 0$ we have

$$2 > \frac{1+\frac{1}{4}\omega}{2\omega}\frac{\rho^2}{\beta^2+\frac{1}{4}a_0}$$

when ω is the at the money implied variance. When plugging the expression of ω, we see it is equivalent to

$$32Y^2(Y-\beta) > (1+2(Y-\beta))\rho^2$$

where

$$Y = \sqrt{\beta^2 + \frac{1}{4}a_0}$$

For example, we see when $\rho = 0$, $a_0 > 0$, this condition is true. Otherwise, there will be some very mild constraints on a_1, a_0, ρ. We want to prove now there are pairs a_1, a_0, ρ so that we have the convexity inequality true everywhere hence guarantee the no arbitrage condition. We reorganize the inequality Equation (52) into

$$2 - 2\frac{x}{\omega}\frac{\rho+\frac{1}{2}x}{\sqrt{\beta^2+1/4\alpha}} + \frac{x^2}{2\omega^2}\frac{(\rho+1/2x)^2}{\beta^2+1/4\alpha} \geq \frac{\frac{1}{4}\omega^2+\omega}{2\omega^2}\frac{(\rho+1/2x)^2}{\beta^2+1/4\alpha}$$

Let

$$\tilde{\beta} = \sqrt{\beta^2 + a_0 - \rho^2}$$

the above inequality is equivalent to

$$\left(\sqrt{2} - \frac{x}{\sqrt{2}\omega}\frac{\rho+\frac{1}{2}x}{\sqrt{\tilde{\beta}^2+\frac{1}{4}(\rho+\frac{1}{2}x)^2}}\right)^2 \geq \frac{\frac{1}{4}\omega+1}{2\omega}\frac{(\rho+\frac{1}{2}x)^2}{\tilde{\beta}^2+\frac{1}{4}(\rho+rac12x)^2}$$

which is further equivalent to

$$\left(2\omega\frac{\sqrt{\tilde{\beta}^2+\frac{1}{4}(\rho+\frac{1}{2}x)^2}}{\rho+\frac{1}{2}x} - x\right)^2 \geq \frac{1}{4}\omega^2+\omega \qquad (53)$$

Theorem 4. For any interval $(-K_1, K_2)$, we can always find $\tilde{\beta}$ and β which are relatively large, so that the parametrization given by

$$a(\omega) = -a_1\omega + a_0 + 1b(\omega) = \omega$$

admits no arbitrage. In particular, the Equations (47) and (53) are both true everywhere for $K \in (-K_1, K_2)$.

Proof. It is easy to spot that when $K \in (-K_1, K_2)$, the parameter d_2 is bounded independent of $\tilde{\beta}$. But we also see that

$$\lim_{\beta\to\infty} -\frac{1}{2\sqrt{\omega}}\frac{\rho+\frac{1}{2}\ln(F/K)}{\sqrt{\beta^2+\frac{1}{4}\alpha}} = 0$$

Therefore the inequality Equation (47) is true in the interval $(-K_1, K_2)$. For the same reason, the inequality Equation (53) is also true. This proves that we can guarantee our remaining variance model is arbitrage free on any finite interval.

CALIBRATION TO MARKETS

We now give some examples. Our data were seven given market-implied volatilities generated on June 28, 2012. We minimized error and compared the calibrated three parameters model vol to the seven given market-implied volatilities. Our preliminary conclusion is that with this three-parameter model, we can fit the seven market-implied volatilities very well. See Exhibits 1 to 14 below.

EXHIBIT 1
Calibration of 1M by 5Y Year Swaption

Forward Rate	Strike	Market Implied	Calculated	SABR
0.80 %	0.10 %	146.23 %	141.51 %	135.20 %
0.80 %	0.20 %	99.51 %	99.42 %	102.29 %
0.80 %	0.30 %	68.70 %	76.65 %	82.21 %
0.80 %	0.60 %	45.84 %	48.51 %	49.89 %
0.80 %	0.70 %	46.53 %	46.31 %	45.96 %
0.80 %	0.80 %	48.48 %	46.19 %	45.73 %
0.80 %	0.80 %	48.48 %	46.19 %	45.73 %
0.80 %	0.90 %	50.56 %	47.49 %	47.89 %
0.80 %	1.00 %	52.62 %	49.67 %	50.90 %
0.80 %	1.30 %	58.12 %	58.50 %	59.66 %
0.80 %	1.50 %	62.00 %	64.80 %	64.39 %
0.80 %	1.60 %	64.25 %	67.89 %	66.46 %

$a_1 = 6.79$, $a_0 = 0.12$, $\rho = -0.02$

Exhibit 2
Calibration of 1M by 10Y Year Swaption

Forward Rate	Strike	Market Implied	Calculated	SABR
1.68 %	0.88 %	51.14 %	50.53 %	50.64 %
1.68 %	0.98 %	47.91 %	48.25 %	48.25 %
1.68 %	1.18 %	44.55 %	45.21 %	44.99 %
1.68 %	1.48 %	43.15 %	43.35 %	43.17 %
1.68 %	1.58 %	43.16 %	43.23 %	43.15 %
1.68 %	1.68 %	43.52 %	43.30 %	43.33 %
1.68 %	1.68 %	43.52 %	43.30 %	43.33 %
1.68 %	1.78 %	43.75 %	43.51 %	43.65 %
1.68 %	1.88 %	44.12 %	43.85 %	44.07 %
1.68 %	2.18 %	45.58 %	45.40 %	45.60 %
1.68 %	2.38 %	46.60 %	46.74 %	46.69 %
1.68 %	2.48 %	47.11 %	47.46 %	47.24 %

$a_1 = 15.75$, $a_0 = 0.24$, $\rho = -0.02$

Exhibit 3
Calibration of 3M by 5Y Year Swaption

Forward Rate	Strike	Market Implied	Calculated	SABR
0.85 %	0.15 %	75.21 %	74.68 %	74.14 %
0.85 %	0.20 %	65.68 %	65.96 %	66.38 %
0.85 %	0.25 %	60.01 %	59.89 %	60.55 %
0.85 %	0.55 %	45.49 %	46.82 %	45.71 %
0.85 %	0.65 %	45.65 %	46.42 %	45.60 %
0.85 %	0.85 %	48.47 %	47.79 %	48.21 %
0.85 %	0.85 %	48.47 %	47.79 %	48.21 %
0.85 %	1.05 %	51.63 %	50.48 %	51.52 %
0.85 %	1.15 %	53.02 %	52.03 %	53.05 %
0.85 %	1.45 %	56.61 %	56.83 %	56.90 %
0.85 %	1.55 %	57.67 %	58.41 %	57.97 %
0.85 %	1.65 %	59.15 %	59.97 %	58.95 %

$a_1 = 5.91$, $a_0 = 0.35$, $\rho = -0.14$

Exhibit 4
Calibration of 3M by 10Y Year Swaption

Forward Rate	Strike	Market Implied	Calculated	SABR
1.73 %	0.93 %	49.14 %	48.97 %	49.10 %
1.73 %	1.03 %	47.50 %	47.54 %	47.52 %
1.73 %	1.13 %	46.29 %	46.44 %	46.32 %
1.73 %	1.43 %	44.41 %	44.59 %	44.45 %
1.73 %	1.53 %	44.22 %	44.34 %	44.25 %
1.73 %	1.73 %	44.29 %	44.19 %	44.23 %
1.73 %	1.73 %	44.29 %	44.19 %	44.23 %
1.73 %	1.93 %	44.53 %	44.41 %	44.54 %
1.73 %	2.03 %	44.74 %	44.62 %	44.77 %
1.73 %	2.33 %	45.59 %	45.54 %	45.59 %
1.73 %	2.43 %	45.91 %	45.92 %	45.88 %
1.73 %	2.53 %	46.13 %	46.32 %	46.18 %

$a_1 = 7.91$, $a_0 = 0.40$, $\rho = -0.01$

Exhibit 5
Calibration of 3Y by 5Y Year Swaption

Forward Rate	Strike	Market Implied	Calculated	SABR
1.92 %	0.22 %	76.55 %	76.10 %	77.43 %
1.92 %	0.32 %	68.97 %	69.10 %	68.89 %
1.92 %	0.42 %	63.92 %	64.19 %	63.28 %
1.92 %	0.92 %	51.22 %	51.40 %	50.18 %
1.92 %	1.42 %	45.61 %	45.87 %	45.43 %
1.92 %	1.92 %	43.01 %	43.06 %	43.33 %
1.92 %	1.92 %	43.01 %	43.06 %	43.33 %
1.92 %	2.42 %	41.82 %	41.63 %	42.30 %
1.92 %	2.92 %	41.33 %	40.99 %	41.76 %
1.92 %	3.42 %	41.20 %	40.83 %	41.45 %
1.92 %	4.42 %	41.40 %	41.32 %	41.17 %
1.92 %	5.42 %	41.82 %	42.35 %	41.06 %

$a_1 = 1.26$, $a_0 = 0.78$, $\rho = 0.28$

Exhibit 6
Calibration of 3Y by 10Y Year Swaption

Forward Rate	Strike	Market Implied	Calculated	SABR
2.50 %	0.25 %	72.19 %	72.30 %	73.19 %
2.50 %	0.50 %	59.72 %	59.53 %	58.93 %
2.50 %	1.00 %	47.90 %	47.85 %	47.04 %
2.50 %	1.50 %	41.95 %	42.01 %	41.53 %
2.50 %	2.00 %	38.45 %	38.59 %	38.50 %
2.50 %	2.50 %	36.41 %	36.49 %	36.72 %
2.50 %	2.50 %	36.41 %	36.49 %	36.72 %
2.50 %	3.00 %	35.21 %	35.20 %	35.62 %
2.50 %	3.50 %	34.55 %	34.46 %	34.92 %
2.50 %	4.00 %	34.26 %	34.09 %	34.47 %
2.50 %	5.00 %	34.15 %	34.05 %	33.95 %
2.50 %	6.00 %	34.39 %	34.56 %	33.70 %

$a_1 = 1.51$, $a_0 = 0.65$, $\rho = 0.30$

Exhibit 7
Calibration of 10Y by 10Y Year Swaption

Forward Rate	Strike	Market Implied	Calculated	SABR
3.15 %	0.65 %	51.59 %	51.23 %	52.28 %
3.15 %	1.15 %	42.36 %	42.37 %	41.88 %
3.15 %	1.65 %	36.44 %	36.66 %	35.93 %
3.15 %	2.15 %	32.38 %	32.61 %	32.04 %
3.15 %	2.65 %	29.50 %	29.64 %	29.37 %
3.15 %	3.15 %	27.41 %	27.46 %	27.52 %
3.15 %	3.15 %	27.41 %	27.46 %	27.52 %
3.15 %	3.65 %	25.98 %	25.91 %	26.24 %
3.15 %	4.15 %	25.04 %	24.84 %	25.38 %
3.15 %	5.15 %	24.11 %	23.79 %	24.42 %
3.15 %	6.65 %	23.99 %	23.94 %	23.94 %
3.15 %	7.65 %	24.37 %	24.67 %	23.89 %

$a_1 = 0.26$, $a_0 = 0.31$, $\rho = 0.30$

EXHIBIT 8
1M by 5Y Implied Volatility

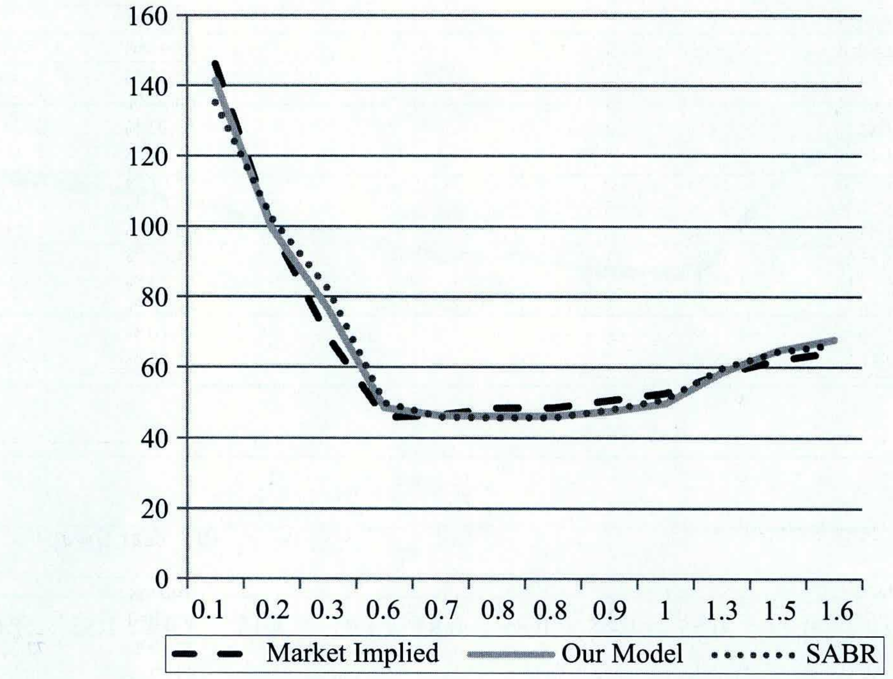


EXHIBIT 9
1M by 10Y Implied Volatility

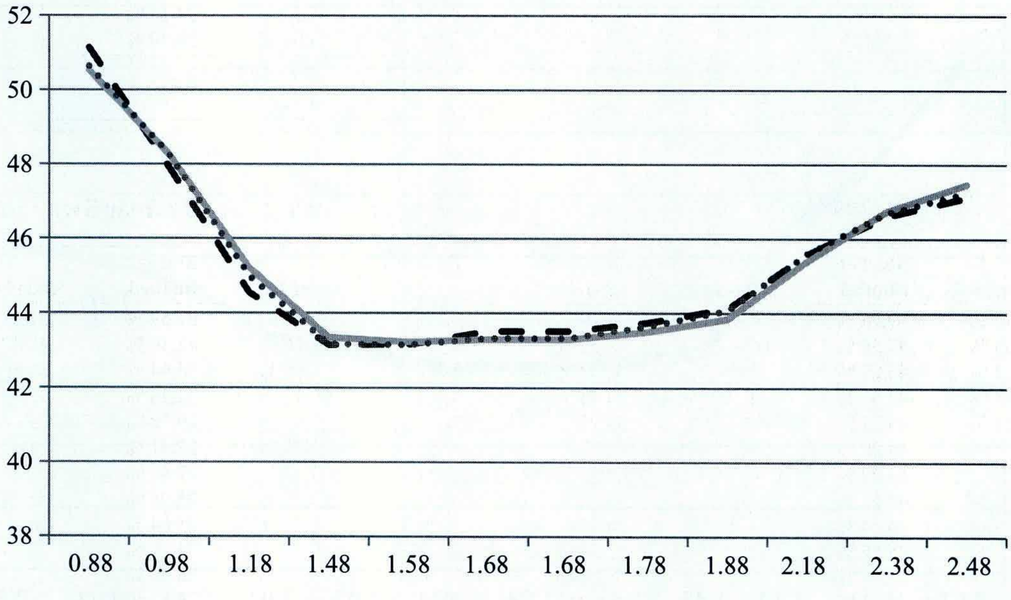


Exhibit 10
3M by 5Y Implied Volatility

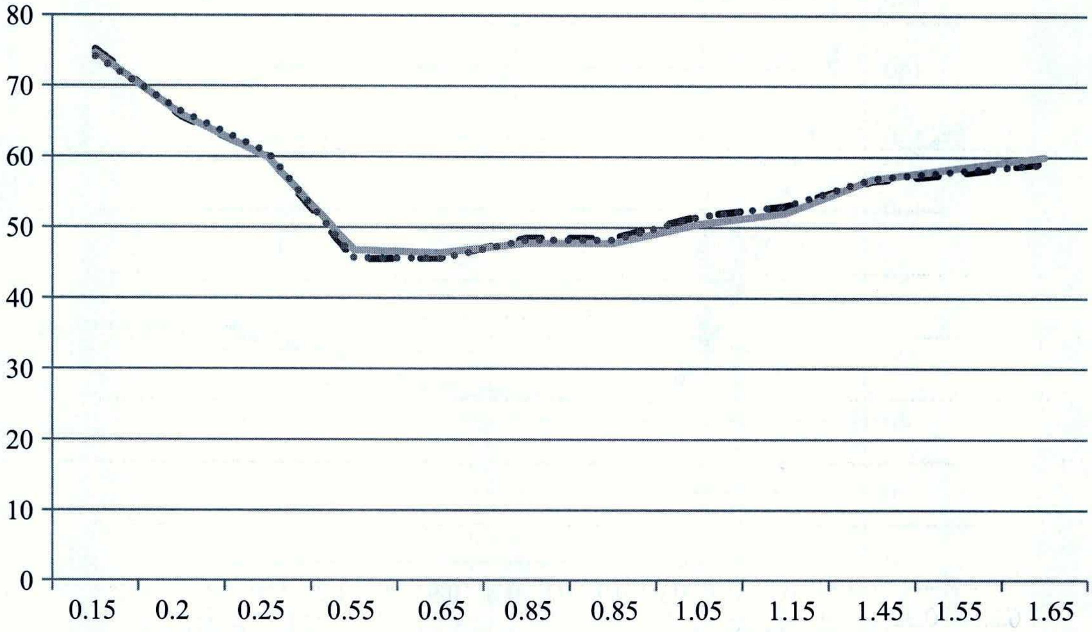


Exhibit 11
3M by 10Y Implied Volatility

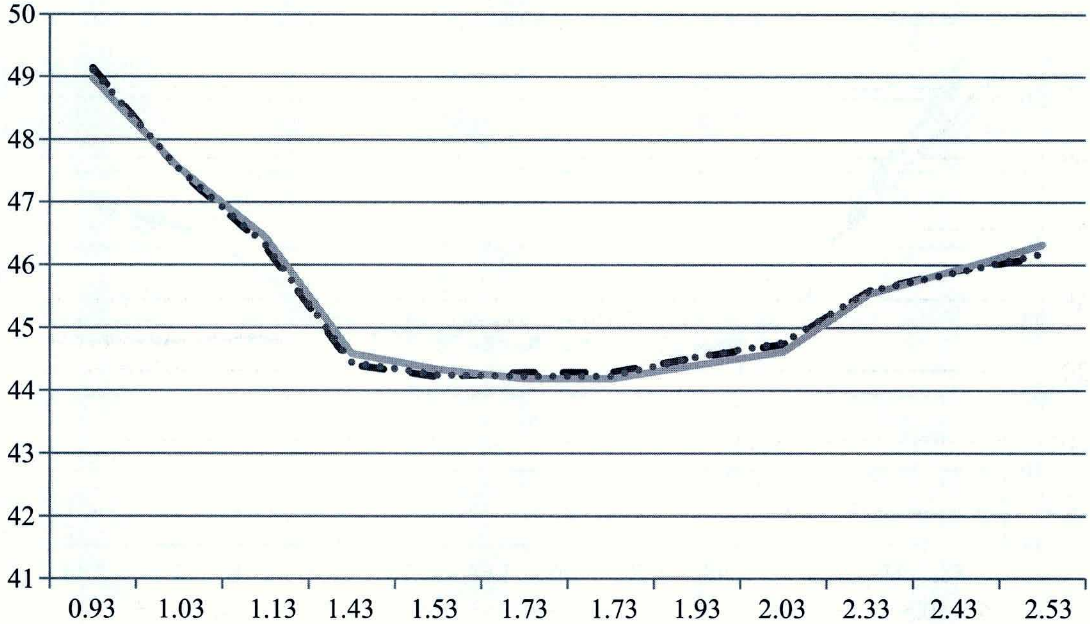


EXHIBIT 12
3Y by 5Y Implied Volatility

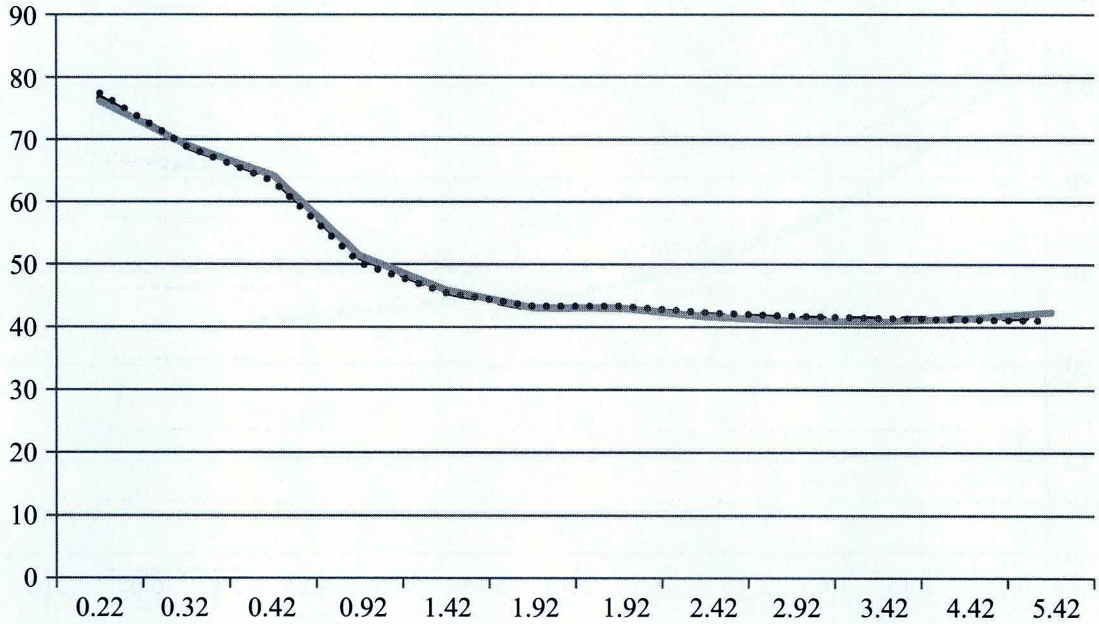


EXHIBIT 13
3Y by 10Y Implied Volatility

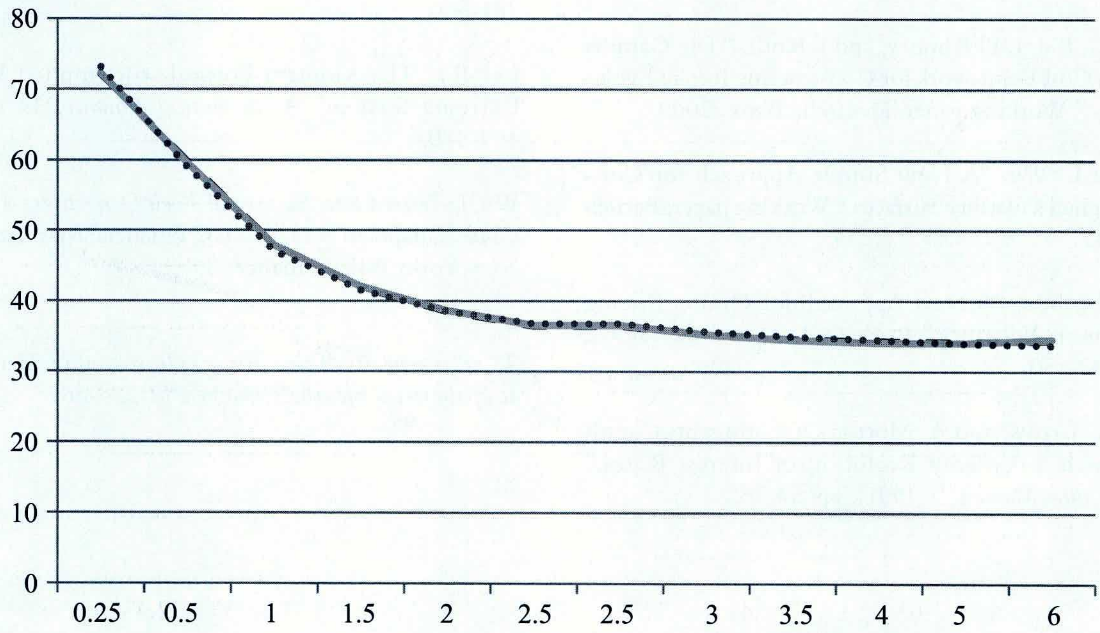
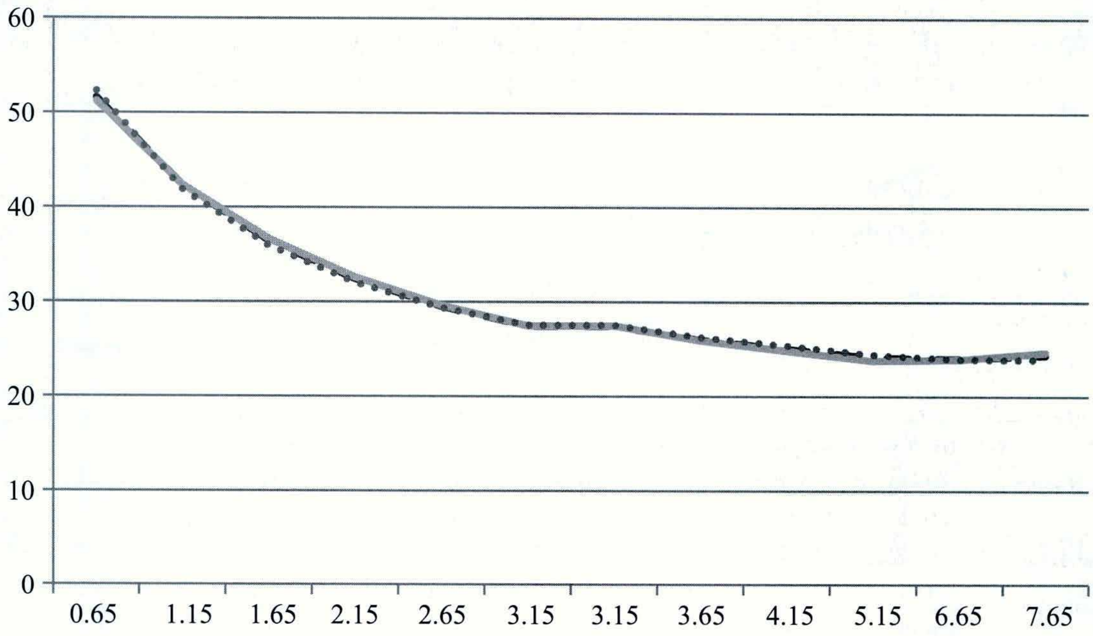


EXHIBIT 14
10Y by 10Y Implied Volatility



ENDNOTE

We thank Travis Fisher and Jason Roth for helpful discussions.

REFERENCES

Arslan M., G. Eid, J. El Khoury, and J. Roth. "The Gamma Vanna Volga Cost Framework for Constructing Implied Volatility Curves." Working paper, Deutsche Bank, 2009.

Carr P., and L. Wu. "A New Simple Approach for Constructing Implied Volatility Surfaces." Working paper, Baruch College, 2012.

Carr P., and J. Sun. "A New Approach for Option Pricing Under Stochastic Volatility." *Review of Derivatives Research*, 2 (2007), pp. 87-150.

Heath D., R. Jarrow, and A. Morton. "Contingents Claims Valuation with a Random Evolution of Interest Rates." *Reviews of Futures Markets*, 9 (1991), pp. 54-76.

Heston, S. "Closed-Form Solution for Options with Stochastic Volatility, with Application to Bond and Currency Options." *Review of Financial Studies*, 6 (1993), pp. 327-343.

Hull J., and A. White. "The Pricing of Options on Assets with Stochastic Volatilities." *Journal of Finance*, 42 (1987), pp. 281-300.

Lee R. "The Moment Formula for Implied Volatility at Extreme Strikes." *Mathematical Finance*, 14 (2004), pp. 469-480.

Wu, L. *Interest Rate Swaps and Their Derivatives: A Practitioner's Guide*. Chapman & Hall/CRC Financial Mathematics Series. New York: Wiley Finance, 2009.