Implied Vol Constraints

by Peter Carr
Bloomberg

Initial version: Sept. 22, 2000

Current version: November 2, 2004

File reference: impvolconstrs3.tex

I am solely responsible for any errors.

I Introduction

This document derives a set of restrictions which implied volatility must satisfy in order to be consistent with both no arbitrage and some mild restrictions on process dynamics. Some of these restrictions on implied volatility were derived previously in Hardy[2] and in Gatheral et. al.[1], but there are also new restrictions here[1]. Any violation of these restrictions implies either the existence of an arbitrage opportunity or a refutation of the process restrictions (or both).

We will proceed as follows. First, for each maturity we formulate our process restrictions by specifying the existence and nature of the risk-neutral density regarded as a function of the future stock price. We also derive the implications of no arbitrage for this density. Then we derive implications for European option values and for implied volatilities across strikes. Second, we derive cross maturity restrictions for the risk-neutral density, option values, and implied volatilities.

As stated above, our starting point for the analysis is a specification of the existence and nature of the risk-neutral density. An alternative starting point could have been a specification of the structure of observed asset prices and trading opportunities. We felt that is it easier to impose restrictions on the density rather than on prices, but our approach is not without its costs in terms of generality. For example, we will show using our approach that the existence of a risk-neutral density implies that call prices decline towards zero as the strike price increases towards infinity. However, a violation of this feature in the market does not necessarily imply an arbitrage opportunity. The existence of an arbitrage hinges on whether or not one can trade calls at all strikes[2]

The restrictions on the process dynamics which we impose are used to both simplify the analysis and financially clear why such masses would exist at any level other than zero. We also impose limited liability of the underlying asset for financial realism, although it is again more mathematically general to allow negative prices. We confine ourselves to a finite time horizon domain so as to allow implied volatilty surfaces which can only be arbitraged over an infinite time horizon[3]. We do not derive the combined effect of our constraints as strikes get arbitrarily large or as strikes approach zero. For a deep exploration of these effects, we refer the reader to Lee[3].

II Cross Strike Restrictions

II-A Risk-Neutral Density

It is well-known that the absence of arbitrage implies the existence of a probability measure Q under which normalized prices of non-dividend paying assets are martingales. We fix a maturity T and use a pure discount bond maturing at T to normalize asset prices[4]. Assuming that the measure is smooth, the existence of the measure implies the existence of a risk-neutral density:

$$q_0(K,T) \equiv \frac{Q\{S_T \in [K, K+dK]\}}{dK}, \quad K \geq 0, T \in [0, \Upsilon], \tag{1}$$

where the nonnegativity of strikes is due to the limited liability assumed of the underlying stock, and the upper bound Υ on maturity arises from technical concerns.

Since q_0 is a probability density, it satisfies a nonnegativity condition:

$$q_0(K,T) \geq 0, \quad K \geq 0, T \in [0, \Upsilon], \tag{2}$$

an integrability condition:

where S_0 is the initial stock price.

The nonnegativity and integrability conditions imply the following Dirichlet upper boundary condition:

$$\lim_{K\uparrow\infty} q_0(K,T) = 0, \quad T \in [0, \Upsilon]. \tag{5}$$

The requirement that the stock absorb at the origin imposes the following Dirichlet lower boundary condition:

$$q_0(0,T) = 0, \quad T \in [0, \Upsilon]. \tag{6}$$

II-B European Option Values

Let $B_0(T) > 0$ be the initial price of a pure discount bond paying one dollar at $T \in [0, \Upsilon]$. At $T = 0$, the bond price is its payoff:

$$B_0(0) = 1. \tag{7}$$

The ability to store money in the future implies that $B_0'(T) \leq 0$, assuming that B_0 is differentiable in T.

Let $F_0(T)$ be the forward price for delivery of one share at date $T \in [0, \Upsilon]$. We demand that the forward measure reprice the forward, and so:

$$F_0(T) \equiv \int_0^\infty L q_0(L,T) dL, \quad T \in [0, \Upsilon]. \tag{8}$$

At $T = 0$, (4) and (8) imply that the forward price is the initial stock price:

$$F_0(0) = S_0. \tag{9}$$

Let $C_0(K,T)$ denote the initial price of a European call struck at $K \geq 0$ and maturing at T. Similarly, We now examine the implications of these definitions for the behavior of call and put values. We focus on call values first. Since $(L-K)^+$ is convex in L, Jensen's inequality and (8) imply that the call is bounded below by its intrinsic value:

$$C_0(K,T) \geq B_0(T)[F_0(T)-K]^+, \quad K \geq 0, T \in [0,\Upsilon]. \tag{12}$$

However, at $T=0$, (4), (7), (9), and (10) imply that the call's price equals its intrinsic value:

$$C_0(K,0) = (S_0-K)^+, \quad K \geq 0. \tag{13}$$

Setting $K=0$ in (10) and using (8) implies that zero strike calls are the product of the discount function and forward prices:

$$C_0(0,T) = B_0(T)F_0(T), \quad T \in [0,\Upsilon]. \tag{14}$$

Setting $K=\infty$ in (10) implies that call values vanish asymptotically:

$$\lim_{K\uparrow\infty} C_0(K,T) = 0, \quad T \in [0,\Upsilon]. \tag{15}$$

Assuming that the initial call prices are differentiable in strike, differentiating (10) w.r.t. K implies that the call slope is proportional to the risk-neutral complementary distribution function for each fixed maturity T:

$$\frac{\partial}{\partial K}C_0(K,T) = -B_0(T)\int_K^\infty q_0(L,T)dL, \quad K \geq 0, T \in [0,\Upsilon]. \tag{16}$$

This condition, the positivity of bond prices, and the non-negativity of the risk-neutral PDF in (2) imply that call values are decreasing in strike:

$$\frac{\partial}{\partial K}C_0(K,T) \leq 0, \quad K \geq 0, \quad T \in [0,\Upsilon]. \tag{17}$$

Since $K=0$ is the lowest possible strike, (17) and (14) imply the following upper bound on a call's value

Setting $K = \infty$ in (17) implies that the call slope vanishes asymptotically:

$$\lim_{K \uparrow \infty} \frac{\partial}{\partial K} C_0(K,T) = 0, \quad T \in [0, \Upsilon]. \tag{20}$$

Differentiating (16) w.r.t. K implies that the call's curvature in strike is the product of the discount function and the risk-neutral density:

$$\frac{\partial^2 C_0}{\partial K^2}(K,T) = B_0(T) q_0(K,T), \quad K \geq 0, T \in [0, \Upsilon]. \tag{21}$$

This condition, the positivity of bond prices, and the non-negativity of the risk-neutral density in (2) implies that call values are convex in strike:

$$\frac{\partial^2 C_0}{\partial K^2}(K,T) \geq 0, \quad K \geq 0, T \in [0, \Upsilon]. \tag{22}$$

At zero strike, the absorbing boundary condition (6) implies that the curvature in strike vanishes:

$$\frac{\partial^2 C_0}{\partial K^2}(0,T) = 0, \quad T \in [0, \Upsilon]. \tag{23}$$

As the strike approaches infinity, (5) and (21) imply that the curvature in strikes also vanishes:

$$\lim_{K \uparrow \infty} \frac{\partial^2 C_0}{\partial K^2}(K,T) = 0, \quad T \in [0, \Upsilon]. \tag{24}$$

Summarizing the results thus far, we have shown that if bond prices are positive, if the forward and call values are defined in terms of the risk-neutral density by (8) and (10) respectively, and if (2) to (6) hold for this density, then one obtains the following necessary conditions on call prices (assuming the requisite smoothness):

$$C_0(K,0) = (S_0 - K)^+, \quad K \geq 0. \tag{25}$$

Conversely, if bond prices are positive, if $q_0(K,T)$ is defined from (21) by:

$$q_0(K,T) \equiv \frac{\frac{\partial^2 C_0}{\partial K^2}(K,T)}{B_0(T)}, \quad K \geq 0, T \in [0, \Upsilon], \tag{29}$$

and if (25) to (28) hold, then it is straightforward to verify that (2) to (6) hold.

We now examine the implications of the definition (11) for the behavior of put values. Since $(K-L)^+$ is convex in L, Jensen's inequality and (8) imply that the put is bounded below by its intrinsic value:

$$P_0(K,T) \geq B_0(T)[K - F_0(T)]^+, \quad K \geq 0, T \in [0, \Upsilon]. \tag{30}$$

However, at $T = 0$, (4), (7), (9), and (11) imply that the put's price equals its intrinsic value:

$$P_0(K,0) = (K - S_0)^+, \quad K \geq 0. \tag{31}$$

Since zero is the lowest possible stock price, (11) implies the following upper bound on a put's value:

$$P_0(K,T) \leq B_0(T)K, \quad K \geq 0, T \in [0, \Upsilon]. \tag{32}$$

Setting $K = 0$ in (11) and using (8) implies that zero strike puts are worthless:

$$P_0(0,T) = 0, \quad T \in [0, \Upsilon]. \tag{33}$$

Setting $K = \infty$ in (11) implies that put values increase without limit:

$$\lim_{K \uparrow \infty} P_0(K,T) = \infty, \quad T \in [0, \Upsilon]. \tag{34}$$

Assuming that the initial put prices are differentiable in strike, differentiating (11) w.r.t. K implies that the put slope is proportional to the risk-neutral distribution function for each fixed maturity T:

Setting $K = 0$ in (35) implies that the put slope vanishes at zero :

$$\frac{\partial}{\partial K} P_0(0, T) = 0, \quad T \in [0, \Upsilon]. \tag{37}$$

Setting $K = \infty$ in (35) and using (3) implies that the put slope is the bond price asymptotically:

$$\lim_{K \uparrow \infty} \frac{\partial}{\partial K} P_0(K, T) = B_0(T), \quad T \in [0, \Upsilon]. \tag{38}$$

Differentiating (35) w.r.t. K implies that the put's curvature in strike is the product of the discount function and the risk-neutral density:

$$\frac{\partial^2 P_0}{\partial K^2}(K, T) = B_0(T) q_0(K, T), \quad K \geq 0, T \in [0, \Upsilon]. \tag{39}$$

This condition, the positivity of bond prices, and the non-negativity of the risk-neutral density in (2) implies that put values are convex in strike:

$$\frac{\partial^2 P_0}{\partial K^2}(K, T) \geq 0, \quad K \geq 0, T \in [0, \Upsilon]. \tag{40}$$

At zero strike, the absorbing boundary condition (6) implies that the curvature in strike vanishes:

$$\frac{\partial^2 P_0}{\partial K^2}(0, T) = 0, \quad T \in [0, \Upsilon]. \tag{41}$$

As the strike approaches infinity, (5) and (39) imply that the curvature in strikes also vanishes:

$$\lim_{K \uparrow \infty} \frac{\partial^2 P_0}{\partial K^2}(K, T) = 0, \quad T \in [0, \Upsilon]. \tag{42}$$

Summarizing the results on put prices, we have shown that if bond prices are positive, if the forward and put values are defined in terms of the risk-neutral density by (8) and (11) respectively, and if (2) to

$$\frac{\partial}{\partial K}P_0(0,T) = 0, \frac{\partial}{\partial K}P_0(K,T) \geq 0, K > 0, \lim_{K\uparrow\infty} \frac{\partial}{\partial K}P_0(K,T) = B_0(T), T \in [0,\Upsilon]. \tag{45}$$

$$\frac{\partial^2}{\partial K^2}P_0(0,T) = 0, \frac{\partial^2}{\partial K^2}P_0(K,T) \geq 0, K > 0, \lim_{K\uparrow\infty} \frac{\partial^2}{\partial K^2}P_0(K,T) = 0, T \in [0,\Upsilon]. \tag{46}$$

Conversely, if bond prices are positive, if $q_0(K,T)$ is defined from (39) by:

$$q_0(K,T) \equiv \frac{\frac{\partial^2 P_0}{\partial K^2}(K,T)}{B_0(T)}, \quad K \geq 0, T \in [0,\Upsilon], \tag{47}$$

and if (43) to (46) hold, then it is straightforward to verify that (2) to (6) hold.

II-C Implied Volatility Surface

Since some of the restrictions to follow will involve derivatives w.r.t. maturity, it will be easier to work with spot prices rather than forward prices. Thus, assuming that bond, stock, and option prices are arbitrage-free, the initial implied volatility surface is defined by:

$$\sigma_{i0}(K,T) \equiv BSC^{-1}(S_0, t_0; K, T; r_{0T}, q_{0T}, C_0(K,T)), \quad K > 0, T \in (0,\Upsilon]. \tag{48}$$

where BSC^{-1} is the inverse of the Black Scholes call formula in volatility, and r_{0T} and q_{0T} are the continuously compounded bond yield and dividend yield over $[0,T]$. Note that we require positive strikes and maturities since implied volatility is not well defined otherwise. Also note that we used the call formula and call prices for definiteness, but under no arbitrage the use of puts must yield the same answer. From (48):

$$C_0(K,T) \equiv BSC(S_0, t_0; K, T; r_{0T}, q_{0T}, \sigma_{i0}(K,T)), \quad K > 0, T \in (0,\Upsilon]. \tag{49}$$

From put call parity, we also have:

$$P_0(K,T) \equiv BSP(S_0, t_0; K, T; r_{0T}, q_{0T}, \sigma_{i0}(K,T)), \quad K > 0, T \in (0,\Upsilon]. \tag{50}$$

and the convexity condition:

$$\frac{\partial^2}{\partial K^2} C_0(K,T) \geq 0, \quad K > 0, T \in (0, \Upsilon]. \tag{52}$$

Similarly, the only conditions on the put price behavior in (43) to (46) which do not necessarily hold for any real and positive implied volatility $\sigma_{i0}(K,T)$ are the slope condition:

$$\frac{\partial}{\partial K} P_0(K,T) \geq 0, \quad K > 0, T \in (0, \Upsilon], \tag{53}$$

and the convexity condition:

$$\frac{\partial^2}{\partial K^2} P_0(K,T) \geq 0, \quad K > 0, T \in (0, \Upsilon]. \tag{54}$$

Assuming that the market put prices are differentiable in K, differentiating (50) w.r.t. K implies:

$$\frac{\partial}{\partial K} P_0(K,T) = \frac{\partial BSP}{\partial K} + \frac{\partial BSP}{\partial \sigma} \frac{\partial \sigma_{i0}}{\partial K}(K,T), \quad K > 0, T \in (0, \Upsilon]. \tag{55}$$

Thus, the nonnegativity of a vertical spread in (53) places a lower bound on the slope in strike of implied:

$$\frac{\partial \sigma_t}{\partial K}(T,K) \geq -\frac{\frac{\partial BSP}{\partial K}}{\frac{\partial BSP}{\partial \sigma}}, \quad K > 0, T \in (0, \Upsilon].$$

Now, by the properties of the Black Scholes formula:

$$\frac{\partial BSP}{\partial K} = e^{-\bar{r}_{0T} T} N(-d_2), \quad K > 0, T \in [0, \Upsilon]. \tag{56}$$

$$\frac{\partial BSP}{\partial \sigma} = K e^{-\bar{r}_{0T} T} \sqrt{T} N'(d_2), \quad K > 0, T \in [0, \Upsilon]. \tag{57}$$

Substituting these results in (55) implies that the lower bound on the strike slope simplifies to:

$$\frac{\partial \sigma_{i0}}{\partial K}(T,K) \geq -\frac{N(-d_2)}{K\sqrt{T} N'(d_2)} \quad K > 0, T \in (0, \Upsilon].$$

Dividing by K implies:

Assuming that the market call prices are twice differentiable in K, differentiating (49) w.r.t. K once implies:

$$\frac{\partial}{\partial K}C_0(K,T) = \frac{\partial BSC}{\partial K} + \frac{\partial BSC}{\partial \sigma}\frac{\partial \sigma_{i0}}{\partial K}(K,T), \quad K>0, T\in(0,\Upsilon]. \tag{59}$$

Thus, the nonpositivity of a vertical spread in (51) places an upper bound on the slope in strike of implied:

$$\frac{\partial \sigma_t}{\partial K}(T,K) \leq -\frac{\frac{\partial BSC}{\partial K}}{\frac{\partial BSC}{\partial \sigma}}, \quad K>0, T\in(0,\Upsilon].$$

Now, by the properties of the Black Scholes formula:

$$\frac{\partial BSC}{\partial K} = -e^{-\bar{r}_{0T}T}N(d_2), \quad K>0, T\in[0,\Upsilon]. \tag{60}$$

$$\frac{\partial BSC}{\partial \sigma} = Ke^{-\bar{r}_{0T}T}\sqrt{T}N'(d_2), \quad K>0, T\in[0,\Upsilon]. \tag{61}$$

Substituting these results in (59) implies that the upper bound on the strike slope simplifies to:

$$\frac{\partial \sigma_{i0}}{\partial K}(T,K) \leq \frac{N(d_2)}{K\sqrt{T}N'(d_2)} \quad K>0, T\in(0,\Upsilon].$$

Dividing by K implies:

$$\frac{\partial \sigma_{i0}}{\partial \ln K}(T,K) \leq \frac{1-N(-d_2)}{\sqrt{T}N'(-d_2)} = \frac{R(-d_2)}{\sqrt{T}}, \quad K>0, T\in[0,\Upsilon], \tag{62}$$

where recall $R(d) \equiv \frac{1-N(d)}{N'(d)}$ is Mill's Ratio. Loosely speaking, (62) says that if the strike is increased by one percent, then implied vol cannot increase by more than Mill's ratio divided by the square root of the time to maturity.

Thus, we have negative lower bounds and positive upper bounds on the slope in strike of implied volatility:

$$-\frac{R(d_2)}{\sqrt{T}} \leq \frac{\partial \sigma_{i0}}{\partial \ln K}(T,K) \leq \frac{R(-d_2)}{\sqrt{T}}, \quad K>0, T\in[0,\Upsilon], \tag{63}$$

Since $d_2 = \frac{\ln(F/K)}{\sigma_{i0}(K,T)\sqrt{T}} - \frac{\sigma_{i0}(K,T)\sqrt{T}}{2}$,

$$\begin{aligned}\lim_{T\uparrow\infty} d_2 &= -\infty \\ \lim_{T\uparrow\infty} -d_2 &= \infty.\end{aligned}$$

Thus, we have:

$$\lim_{T\uparrow\infty} -\frac{R(-\infty)}{\sqrt{T}} \leq \lim_{T\uparrow\infty} \frac{\partial\sigma_{i0}}{\partial \ln K}(T,K) \leq \lim_{T\uparrow\infty} \frac{R(\infty)}{\sqrt{T}}, \quad K > 0, T \in [0, \Upsilon]. \tag{64}$$

To determine $R(-\infty)$ and $R(\infty)$, recall the definition of Mill's ratio:

$$R(d) \equiv \frac{1 - N(d)}{N'(d)}. \tag{65}$$

Using L'Hopital's rule:

$$\begin{aligned}\lim_{d\uparrow\infty} R(d) &= \lim_{d\uparrow\infty} \frac{1}{d} = 0 \\ \lim_{-d\downarrow\infty} R(d) &= \lim_{-d\downarrow\infty} \frac{1}{d} = 0.\end{aligned}$$

Substituting this result in (64) implies that the slope in strike of the implied vanishes asymptotically:

$$\lim_{T\uparrow\infty} \frac{\partial\sigma_{i0}}{\partial \ln K}(T,K) = 0.$$

Furthermore, the leading order in the rate of decline is $O\left(\frac{1}{T}\right)$. If implieds are graphed against $d \equiv \frac{\ln(K/F)}{\sqrt{T}}$, then the leading order in the rate of decline is $O\left(\frac{1}{\sqrt{T}}\right)$. Given this behavior, it is interesting to note that the empirical consistency of the implied slope in the variable d which is observed for the first few years in the S&P market, cannot persist at indefinitely long maturities. The volatility smile must flatten out asymptotically.

Thus, the non-negativity of a butterfly spread in (52) places a bound on the curvature in strike of the initial implied volatility surface:

$$\frac{\partial^2 \sigma_0}{\partial K^2}(K,T) \geq -\frac{\frac{\partial^2 BSC}{\partial K^2}}{\frac{\partial BSC}{\partial \sigma}} - 2\frac{\frac{\partial^2 BSC}{\partial \sigma \partial K}}{\frac{\partial BSC}{\partial \sigma}}\frac{\partial \sigma_{i0}}{\partial K}(K,T) - \frac{\frac{\partial^2 BSC}{\partial \sigma^2}}{\frac{\partial BSC}{\partial \sigma}}\left(\frac{\partial \sigma_{i0}}{\partial K}(K,T)\right)^2, \quad K > 0, T \in (0, \Upsilon].$$

Now, by the properties of the Black Scholes call formula:

$$\frac{\partial^2 BSC}{\partial K^2} = \frac{e^{-\bar{r}_{0T}T}N'(d_2)}{K\sigma\sqrt{T}}, \quad K > 0, T \in (0, \Upsilon]. \tag{67}$$

$$\frac{\partial^2 BSC}{\partial \sigma \partial K} = e^{-\bar{r}_{0T}T}N'(d_2)\frac{d_1}{\sigma}, \quad K > 0, T \in (0, \Upsilon]. \tag{68}$$

$$\frac{\partial^2 BSC}{\partial \sigma^2} = N'(d_2)\frac{Ke^{-\bar{r}_{0T}T}\sqrt{T}d_1 d_2}{\sigma}, \quad K > 0, T \in (0, \Upsilon]. \tag{69}$$

Substituting these results in (66) implies that the lower bound on curvature simplifies to:

$$\frac{\partial^2 \sigma_{i0}}{\partial K^2}(K,T) \geq -\frac{1}{K^2\sigma_{i0}(K,T)} - \frac{2d_1}{\sigma_{i0}(T,K)K\sqrt{T}}\frac{\partial \sigma_{i0}}{\partial K}(K,T) - \frac{d_1 d_2}{\sigma_{i0}(T,K)}\left(\frac{\partial \sigma_{i0}}{\partial K}(K,T)\right)^2, \quad K > 0, T \in (0, \Upsilon]. \tag{70}$$

Loosely speaking, if the strike is increased by $1, the change in the strike slope of the implied volatility is bounded below. Since puts and calls have the same second strike derivative, the imposition of (54) does not produce any more constraints on the behavior of the implied volatility.

Conversely, if the three cross-strike arbitrage restrictions (58),(62), and (70) hold for the implied volatility, then the corresponding cross-strike conditions (25) to (28) hold for call values and the cross-strike conditions (43) to (46) hold for put values.

III Cross Maturity Restrictions

III-A Risk-neutral Density

denote the density governing the risk-neutral probability that the stock price at T' is near K, conditional on the information set $\mathcal{F}_T$ at the future time T. This information set includes at least bond, stock, and call prices at time T. Since $q_T(K,T')$ is a probability density, it is non-negative:

$$q_T(K,T') \geq 0, \quad K \geq 0, T \in [0,\Upsilon], T' \in [T,\Upsilon], \tag{72}$$

integrates to one:

$$\int_0^\infty q_T(L,T')dL = 1, \quad K \geq 0, T \in [0,\Upsilon], T' \in [T,\Upsilon]. \tag{73}$$

and satisfies the initial condition:

$$q_T(K,T) = \delta(K - S_T), \tag{74}$$

where S_T is the stock price at T. The nonnegativity and integrability conditions imply the following Dirichlet upper boundary condition:

$$\lim_{K\uparrow\infty} q_T(K,T') = 0, \quad T \in [0,\Upsilon], T' \in [T,\Upsilon]. \tag{75}$$

The requirement that the stock absorb at the origin imposes the following Dirichlet lower boundary condition:

$$q_T(0,T') = 0, \quad T \in [0,\Upsilon], T' \in [T,\Upsilon]. \tag{76}$$

III-B Call Prices

The ability to trade in the stock between T and T' implies that under the forward measure, the forward price of maturity T' has zero expected change over $[T,T']$, or equivalently, $q_T(L,T')$ obeys the following "forward repricing condition":

Repeating the unconditional analysis of the last section implies that:

$$C_T(K,T) = (S_T - K)^+, \quad K \geq 0, T \in [0,\Upsilon]. \tag{79}$$

$$C_T(0,T') = B_T(T')F_T(T'), \quad \lim_{K\uparrow\infty} C_T(K,T') = 0, \quad T \in [0,\Upsilon], T' \in [T,\Upsilon]. \tag{80}$$

$$C_T(K,T') \in [B_T(T')(F_T(T') - K)^+, B_T(T')F_T(T')], \quad K > 0, T \in [0,\Upsilon], T' \in [T,\Upsilon]. \tag{81}$$

$$\frac{\partial}{\partial K}C_T(0,T') = B_T(T'), \frac{\partial}{\partial K}C_T(K,T') \leq 0, K > 0, \lim_{K\uparrow\infty}\frac{\partial}{\partial K}C_T(K,T') = 0, T \in [0,\Upsilon], T' \in [T,\Upsilon]. \tag{82}$$

$$\frac{\partial^2}{\partial K^2}C_T(0,T') = 0, \frac{\partial^2}{\partial K^2}C_T(K,T') \geq 0, K > 0, \lim_{K\uparrow\infty}\frac{\partial^2}{\partial K^2}C_T(K,T') = 0, T \in [0,\Upsilon], T' \in [T,\Upsilon]. \tag{83}$$

Conversely, if $q_T(K,T')$ is defined by:

$$q_T(K,T') \equiv \frac{\frac{\partial^2 C_T}{\partial K^2}(K,T')}{B_T(T')}, \quad K \geq 0, T \in [0,\Upsilon], T' \in [T,\Upsilon] \tag{84}$$

and if (79) to (83) hold, then it is straightforward to verify that (72) to (76) hold.

Under deterministic interest rates and dividends, the lower bound in (81) implies that the conditional call value has non-negative time value:

$$C_T(K,T') \geq [S_T e^{-\bar{q}_{T,T'}\tau} - K e^{-\bar{r}_{T,T'}\tau}]^+ = e^{-\bar{q}_{T,T'}\tau}[S_T - K e^{-(\bar{r}_{T,T'} - \bar{q}_{T,T'})\tau}]^+, \quad K > 0, T \in [0,\Upsilon], T' \in [T,\Upsilon]. \tag{85}$$

Taking unconditional risk-neutral expectations implies that a modified calendar spread has non-negative value:

$$C_0(K,T') - e^{-\bar{q}_{T,T'}\tau}C_0(K e^{-(\bar{r}_{T,T'} - \bar{q}_{T,T'})\tau}, T) \geq 0, \quad K \geq 0, T \in [0,\Upsilon], T' \in [T,\Upsilon]. \tag{86}$$

Dividing by $T' - T > 0$, the LHS can be rewritten as:

$$\frac{C_0(K,T') - C_0(K,T)}{T' - T} + [r(T) - q(T)]K\frac{C_0(K,T) - C_0(K e^{-(\bar{r}_{T,T'} - \bar{q}_{T,T'})\tau}, T)}{[r(T) - q(T)]K(T' - T)}$$

Substituting (29) in (88) and simplifying yields the following constraint on the initial risk-neutral density:

$$\int_K^\infty (L-K)\frac{\partial}{\partial T}q_0(L,T)dL + [q(T)-r(T)]\int_K^\infty Lq_0(L,T)dL \geq 0, \quad K \geq 0, T \in [0,\Upsilon]. \tag{89}$$

A corresponding analysis for puts would imply:

$$\frac{\partial}{\partial T}P_0(K,T) + [r(T)-q(T)]K\frac{\partial}{\partial K}P_0(K,T) + q(T)P_0(K,T) \geq 0, \quad K \geq 0, T \in [0,\Upsilon]. \tag{90}$$

Substituting (47) in (90) and simplifying yields the following constraint on the initial risk-neutral density:

$$\int_0^K (K-L)\frac{\partial}{\partial T}q_0(L,T)dL + [q(T)-r(T)]\int_0^K Lq_0(L,T)dL \geq 0, \quad K \geq 0, T \in [0,\Upsilon]. \tag{91}$$

Since (90) can be derived from (88) and put call parity, any risk-neutral density satisfying (88) and put call parity will also satisfy (91).

III-C Implied Volatility

Assuming that the market price of a call is always differentiable in T, differentiating (49) w.r.t. T implies:

$$\frac{\partial}{\partial T}C_0(K,T) = \frac{\partial BSC}{\partial T} + \frac{\partial BSC}{\partial \sigma}\frac{\partial \sigma_{i0}}{\partial T}(K,T), \quad K > 0, T \in (0,\Upsilon]. \tag{92}$$

Thus, the non-negativity of the modified calendar spread value in (88) translates into the following restriction on the initial implied:

$$\frac{\partial BSC}{\partial T} + \frac{\partial BSC}{\partial \sigma}\frac{\partial \sigma_{i0}}{\partial T}(K,T) + [r(T)-q(T)]\left[\frac{\partial BSC}{\partial K} + \frac{\partial BSC}{\partial \sigma}\frac{\partial \sigma_{i0}}{\partial K}(K,T)\right] + q(T)BSC \geq 0, \quad K > 0, T \in (0,\Upsilon],$$

or equivalently:

$$\sigma_{i0}^2(K,T)K^2\,\partial^2 BSC \;+\; \partial BSC\,\partial\sigma_{i0}(K,T) + [r(T) - q(T)]\,\partial BSC\,\partial\sigma_{i0}(K,T) \geq 0, \quad K > 0, T \in (0,\Upsilon]$$

by the properties (61) and (67) of the Black Scholes formula. Dividing both sides by $\sigma_{i0}(K,T)$ implies:

$$\frac{\partial \ln \sigma_{i0}}{\partial T}(K,T) \geq -\frac{1}{2} - [r(T) - q(T)]\frac{\partial \ln \sigma_{i0}}{\partial K}(K,T), \quad K > 0, T \in (0, \Upsilon]. \tag{93}$$

Loosely speaking, if $r(T) = q(T)$ and the maturity is increased slightly, then implied volatility cannot fall by more than half a percent of its former value.

Conversely, if the cross-maturity arbitrage restriction (93) holds for the initial implied volatility surface, then the corresponding cross-maturity condition (88) holds for the initial call values.

Recall the arbitrage restrictions across strike on implied volatility:

$$\frac{R(d_2)}{\sqrt{T}} \leq \frac{\partial \sigma_{i0}}{\partial \ln K}(T,K) \leq \frac{R(-d_2)}{\sqrt{T}}, \quad K > 0, T \in [0, \Upsilon], \tag{94}$$

where $R(d) \equiv \frac{1-N(d)}{N'(d)}$ is Mill's Ratio, and:

$$\frac{\partial^2 \sigma_{i0}}{\partial K^2}(K,T) \geq -\frac{1}{K^2 \sigma_{i0}(K,T)} - \frac{2d_1}{\sigma_{i0}(T,K)K\sqrt{T}}\frac{\partial \sigma_{i0}}{\partial K}(K,T)\frac{d_1 d_2}{\sigma_{i0}(T,K)}\left(\frac{\partial \sigma_{i0}}{\partial K}(K,T)\right)^2, \quad K > 0, T \in (0, \Upsilon]. \tag{95}$$

Taken together, the three restrictions (93),(94), and (95) imply that when implied volatility is plotted against strike and maturity, no arbitrage requires that it cannot fall too fast with maturity nor change too much with strike. Furthermore, while the implied volatility can be concave in strike, it cannot be too concave.

Conversely, if the three restrictions on the implied volatility surface hold, then if the call and put prices at strike K and maturity T are quoted by using the appropriate Black Scholes formula with the corresponding implied vol as in (49), then these quotes are arbitrage-free.

Bibliography

[1] Gatheral J., A. Matytsin, and C. Youssfi, 2000, "Rational Shapes of the Volatility Surface", Merrill Lynch presentation at the Columbia Practitioner's Conference, New York, October 2000.

[2] Hodges, H., 1996, "Arbitrage Bounds on the Implied Volatility Strike and Term Structures of European-Style Options", *The Journal of Derivatives*, Summer 1996, 23–35.

[3] Lee, R., 2004, "The Moment Formula for Implied Volatility at Extreme Strikes", forthcoming in *Mathematical Finance*.

[4] Patel, J. and C. Read, 1996, Handbook of the Normal Distribution, Second Edition, Marcel Dekker, Inc, New York.