Vol, Skew, and Smile Trading

Peter Carr and Liuren Wu
pcfarr@nyc.rr.com and Liuren.Wu@baruch.cuny.edu

Abstract

In general, an option’s fair value depends crucially on the volatility of its underlying asset. In a stochastic volatility (SV) setting, an at-the-money straddle can be dynamically traded to profit on average from the difference between its underlying’s instantaneous variance rate and its Black Merton Scholes (BMS) implied variance rate. In SV models, an option’s fair value also depends on the covariation rate between returns and volatility. We show that a pair of out-of-the-money options can be dynamically traded to profit on average from the difference between this instantaneous covariation rate and half the slope of a BMS implied variance curve. Finally, in SV models, an option’s fair value also depends on the variance rate of volatility. We show that an option triple can be dynamically traded to profit on average from the difference between this instantaneous variance rate and a convexity measure of the BMS implied variance curve. Our results yield precise financial interpretations of particular measures of the level, slope, and curvature of a BMS implied variance curve. These interpretations help explain standard quotation conventions found in the over-the-counter market for options written on precious metals and on foreign exchange.

We are grateful to Zameer Arora, Josh Birnbaum, Sebastien Bossu, Bruno Dupire, Travis Fisher, Apollo Hogan, Chou Liang, Sergey Nadtochiy, Jeremy Primer, Jian Sun, Pratik Worah, Zhibai Zhang, and Jim Zhu for comments. They are not responsible for any errors.
1 Introduction

In general, an option’s fair value depends crucially on the volatility of its underlying asset. Black Scholes [1] and Merton[4] develop this insight in the simplest possible model, where the volatility of the option’s underlying asset is constant. While this particular assumption of the Black Merton Scholes (BMS) model is never relied upon in practice, the BMS model is still used to define the widely used concept of implied volatility (IV).

In a stochastic volatility (SV) setting, it is well known that the realized variance of an option’s underlying asset can be robustly replicated by combining a static option in co-terminal options across all strike prices with dynamic trading in the underlying asset. A measure of the realized covariance between the price relative and the instantaneous variance can also be synthesized by this kind of trading strategy.

These theoretical advances have lead to practical consequences such as the rise of variance and gamma swaps and the 2003 redesign of the Volatility Index (VIX). Nonetheless, there remain markets where other methods for trading variance and covariance dominate. In particular, the over-the counter (OTC) options market for foreign exchange (FX) provides transparency in at most five strike rates at each maturity. Perhaps as a result, there is no analog of the VIX for FX volatility. We show that in a partially unspecified SV setting, an at-the-money (ATM) straddle can be dynamically traded to profit on average from the difference between its underlying’s instantaneous variance rate and its BMS implied variance rate. In SV models, an option’s fair value also depends on the covariation rate between returns and volatility. We show that a pair of out-of-the-money (OTM) options can be dynamically traded to profit on average from the difference between this instantaneous covariation rate and half the slope of a line connecting two points on a BMS implied variance curve. Finally, in SV models, an option’s fair value also depends on the variance rate of volatility. We show that an option triple can be dynamically traded to profit on average from the difference between this instantaneous variance rate and a standard convexity measure of the BMS implied variance curve. Our results yield precise financial interpretations of particular measures of the level, slope, and curvature of a BMS implied variance curve. These interpretations help explain standard quotation conventions found in the OTC market for options written on precious metals and on FX.

As is well known, an FX option gives its owner the right to exchange one currency for another. Let $\mathbb{P}$ be the real world probability measure used to define an arbitrage opportunity in an FX market. We suppose that neither currency can become worthless relative to the other. From the fundamental theorem of asset pricing, a consequence of no currency arbitrage is the existence of a pair of probability measures, $Q_-$ and $Q_+$, which are each equivalent to $\mathbb{P}$. $Q_-$ is the probability measure arising when the so-called domestic currency is taken as numeraire, while $Q_+$ is the probability measure arising when the so-called foreign currency is taken as numeraire. Let $S$ be the spot FX rate, expressed in domestic currency units per foreign currency unit. Assuming zero interest rates, $Q_-$ is called an equivalent martingale measure because $S$ is a (strictly positive) local martingale under $Q_-$. The negative sign subscripting $Q_-$ reflects the fact that $\ln S$ has negative drift under $Q_-$. $Q_+$ is also called an equivalent martingale measure because the reciprocal FX rate $R = \frac{1}{S}$ is a (strictly positive) local martingale under $Q_+$. The positive sign subscripting $Q_+$ reflects
the fact that \( \ln S = -\ln R \) has positive drift under \( Q_+ \).

The dynamics of a European option’s value under either one of these probability measures are determined by the corresponding dynamics of both the underlying FX rate and its BMS IV. In this paper, we assume that both of these stochastic processes evolve under \( Q_- \) as a strictly positive continuous local martingale. We consider an FX options market-maker who quotes IV by strike rate for all positive strike rates and for some fixed maturity date. We further suppose that as calendar time evolves, the market-maker continuously updates all of his IV quotes by the same random percentage amount. In this setting, the instantaneous risk-neutral drift in the value of a general option portfolio depends on three stochastic processes, namely the instantaneous variance rate of the underlying log FX rate, the instantaneous covariation of this log FX rate with the log IV curve, and the instantaneous variance rate of the log IV curve. We develop three special option portfolios which allow an investor to synthesize a short term forward contract written on exactly one of these three constructs, despite a complete lack of knowledge regarding the level and dynamics of the other two. These three special option portfolios resemble the three portfolios which liquidly trade in the OTC options markets for precious metals and for FX. Furthermore, the vegas of these three special option portfolios are positively proportional to the three standard quotes found in these options market, which are positive multiples of the level, slope, and curvature of an IV curve.

In this paper, we develop a hypothetical options market similar in structure to the OTC FX options market. The OTC FX options market is the deepest, largest and most liquid market for options of any kind. In the OTC FX options market, dealers stand ready to trade an at-the-money (ATM) straddle, a risk-reversal (risky), and/or a butterfly-spread (fly). As is well known, a straddle combines one put and one call with the same underlying, strike rate, and maturity date. In the OTC FX options market, the straddle is said to be ATM when the common strike rate leads to zero delta for the straddle. As with all of the greeks used in the OTC FX options market, the straddle delta is calculated by inputting the quoted IV into the relevant BMS formula.

The resulting strike rate for the ATM straddle is almost never equal to the forward FX rate. Puts are said to be out-of-the-money (OTM) when their strike rates are below this ATM strike, while calls are said to be OTM when their strike rates are above. In contrast to a straddle which has only long positions, a risk-reversal is long one OTM call and short one OTM put. The call and put are required to have the same (BMS) delta up to sign. Clearly, the risk-reversal has positive delta equal to twice the delta of the call component.

In contrast to a risk-reversal whose option positions differ in sign, a strangle combines a long position in one OTM call with a long position in one OTM put with the same delta up to sign. Clearly, a strangle has positive (BMS) vega. A butterfly-spread is long one strangle and short one ATM straddle.

In the OTC FX options market, the three terms ATM straddle, risk-reversal, and butterfly-spread ambiguously refer not only to the three option portfolios described above, but also to three quotes. Besides describing a long position in a call and put whose delta vanishes, the term ATM straddle is also the common IV that the call and put possess. Besides describing a long OTM call and short put position, the term risk-reversal is also the difference between the IV of the OTM call and the IV of the OTM put. Finally, besides describing a long OTM strangle and short ATM straddle position, the term butterfly-spread also refers to the difference between the average of two
equally OTM IV’s and the ATM IV. Clearly, these three standard FX option quotes are positive multiples of the level, slope, and curvature of the IV curve.

The main objective of this paper is to develop the informational content of the level, slope, and curvature of an implied variance curve. We develop a particular measure of moneyness such that the ATM implied variance represents the payment leg of an impure bet on the variance rate of log FX. We also develop a particular measure of the implied variance slope (called skew) which represents the payment leg of an impure bet on the covariation rate of log FX with log IV. Finally, we develop a particular measure of the implied variance curvature (called smile) which represents the payment leg of an impure bet on the variance rate of IV.

Our ATM implied variance, skew, and smile all represent the payment leg of a short term forward contract written on either a variance rate or a covariation rate. The forward contract is synthesized up to zero mean error by dynamic trading in either an ATM straddle, a normalized risk-reversal, or a normalized butterfly-spread. We refer to these three dynamic trading strategies as the vol trade, the skew trade, and the smile trade respectively. The zero mean error arises from non-vanishing delta and/or vega in these three special trades. The forward contract replication becomes perfect when an investor can both delta and vega hedge without incurring a cost. The absence of arbitrage in portfolios involving ATM straddles implies in our setting that the ATM IV always equals the instantaneous volatility. When this occurs, the ATM straddle value has zero risk-neutral drift and vega hedging becomes costless. However in this paper, we are interested in knowing when skew and smile trades are riskless arbitrages without having to assume that there is no arbitrage in the ATM straddle quote. We give sufficient conditions under which a market maker’s IV quotes cause skew and smile trades to be riskless arbitrages.

A byproduct of our analysis is an enhanced understanding of the three positions and the three quotes simultaneously called ATM straddle, risk-reversal, and butterfly-spread in the OTC FX options market. We develop an SV setting where three option portfolios called normalized ATM straddles, normalized risk-reversals, and normalized butterfly-spreads arise naturally. The normalization arises by dividing each option position in the standard OTC FX options definition by a partial derivative that we call relative gamma. The vegas of these three normalized option portfolios then become positively proportional to the three standard FX option quotes describing the level, slope, and curvature of the IV curve.

Our three normalized option portfolios allow investors to bet on either the instantaneous variance rate of the underlying’s log FX rate, the instantaneous covariation rate of the log FX rate with log IV, or the instantaneous variance rate of log IV respectively. The investor is able to bet on exactly one of these three stochastic processes, despite a complete lack of knowledge of either the level or dynamics of the other two. Following standard terminology, we refer to the three bets as vol, skew, and smile trades respectively.

In the OTC FX options market, the standard quotation convention is to quote BMS IV at a few discrete delta levels and for a few discrete terms. In our similar options market, a market-maker continuously quotes an entire IV curve at one fixed maturity date instead. In our fictitious options market, the market-maker quotes IV by strike rate, rather than by BMS delta. However, we will come to understand how the curious delta quoting convention would naturally arise.

There are two contributions of this paper. First, we provide a formal definition of statistical
arbitrage in the context of an SV model where option quotes may allow riskless arbitrage and the only observables are spot FX and the IV curve. In particular, the instantaneous volatilities of spot and IV are not observed and are formally treated as random variables whose distribution is unknown. In this context, statistical arbitrage is defined as positive drift in an option portfolio using any equivalent martingale measure arising from excluding arbitrage between the two currencies. We develop three sufficient alternative conditions under which statistical arbitrage arises.

The second contribution of this paper arises in the slightly more classical context where the instantaneous volatility of spot, $\sigma_t$, is treated as observable and non-random at time $t$. In this case, we give two alternative sufficient conditions under which riskless arbitrage arises. The first of these riskless arbitrages arises when the market-maker sets insufficient slope in his implied variance curve, while the second one arises when the market maker sets insufficient convexity.

An overview of this paper is as follows. In the next section, we do a short literature review on trading spot volatility, covariation of spot and volatility, and volatility of volatility. In the next section, we review valuation formulas, moneyness measures, and relationships between partial derivatives of option values (henceforth greeks) in the zero rates BMS model, which are of relevance in the sequel. In the following section, we state the dynamical restrictions in our SV setting. The following section develops expressions for the instantaneous gain for both a single option position with a single strike rate, eg. an OTM put, and for a portfolio of options with up to three strike rates. In general, we find that the mean gain depends on three stochastic processes, namely the instantaneous variance rates of log FX and log IV, as well as the covariation rate between log FX and log IV. The following section develops three special option portfolios for which the mean gain just depends on one of these three stochastic processes. The next section presents sufficient conditions on the IV curve under which the mean gain on each special option portfolio is positive. When these conditions hold, the special option portfolio is said to be a statistical arbitrage. The penultimate section presents more complicated sufficient conditions on the IV curve under which the mean gain on each special option portfolio is both positive and riskless. When these conditions hold, the special option portfolio is said to be a riskless arbitrage (in our SV model). The final section summarizes the paper and presents extensions for future research.

2 Literature Review

Nadtochiy and Obloj[5] consider trading skew when spot hits a fixed spatial level.

3 BMS Model

Let $S > 0$ be the underlying spot foreign exchange (henceforth FX) rate, expressed in domestic currency units per foreign currency unit. As is well known, $S$ follows geometric Brownian motion in the BMS model. We assume that the domestic interest rate and the foreign interest rate both vanish, so that $S$ has zero risk-neutral drift. Let $\sigma > 0$ be the assumed constant volatility of spot FX in the BMS model.
3.1 BMS Pricing Formulas

We first recall the BMS pricing formula for a European put. Accordingly, let \( P_b \) denote the put price in the zero rates BMS model. The put premium is assumed to be expressed in domestic currency. One put with strike rate \( K > 0 \) and maturity date \( T > 0 \) gives its owner the right to exchange one foreign currency unit for \( K \) domestic currency units at \( T \). Let \( t \in [0, T] \) denote the moving calendar time. Since the BMS model is time-homogeneous, the option premium depends on \( T \) and \( t \) only through their difference \( \tau \equiv T - t \geq 0 \), called the term of the option.

The BMS model value of the put with fixed strike rate \( K \) and fixed maturity date \( T \) is given by a strictly positive function \( P^b(S, \sigma, \tau) : \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+ \mapsto \mathbb{R}^+ \) of three strictly positive variables. In the actual BMS model, \( \sigma \) is a parameter, not a variable. Our treatment of \( \sigma \) as a variable reflects our intended use of the put pricing function \( P_b \) in the next section, where volatility becomes a stochastic process. The zero rates BMS model put pricing function is defined as:

\[
P^b(S, \sigma, \tau) \equiv K N(z_-(K/S, \sigma \sqrt{\tau})) - S N(z_+(K/S, \sigma \sqrt{\tau})), \quad K > 0, \tau \in [0, T],
\]  

where for \( z \in \mathbb{R} \), \( N(z) = \int_{-\infty}^{z} \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy \) is the standard normal cumulative distribution function, whose arguments are defined as:

\[
z_\pm(K/S, \sigma \sqrt{\tau}) \equiv \frac{\ell_\pm(K/S, \sigma \sqrt{\tau})}{\sigma \sqrt{\tau}}, \quad \ell_\pm(K/S, \sigma \sqrt{\tau}) \equiv \ln(K/S) \mp \sigma^2/2.
\]  

From put call parity, the BMS call value function \( C^b(S, \sigma, T - t) \) solves:

\[
C^b(S, \sigma, T - t) = S - K + P^b(S, \sigma, T - t), \quad S > 0, \sigma > 0, T - t > 0.
\]  

Substituting (1) in (3) implies that the BMS call value with fixed strike rate \( K \) and fixed maturity date \( T \) is given by the following strictly positive function \( C^b(S, \sigma, \tau) : \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+ \mapsto \mathbb{R}^+ \):

\[
C^b(S, \sigma, \tau) \equiv S N(-z_-(K/S, \sigma \sqrt{\tau})) - K N(-z_+(K/S, \sigma \sqrt{\tau})), \quad K > 0, \tau \in [0, T].
\]

3.2 BMS Log Moneyness Measures

In this subsection, we develop four measures of log moneyness which will all be used in the sequel. We think of \( \ell_\pm \) defined in (2) as the first two log moneyness measures. To understand first why \( \ell_- \) is a log moneyness measure, let \( Q^b_- \) be the equivalent martingale measure arising in the zero rates BMS model when the domestic currency is taken to be the numeraire. Under this probability measure, the log FX rate has negative drift equal to \(-\sigma^2/2\). As a result:

\[
\ell_-(K/S, \sigma \sqrt{\tau}) \equiv \ln(K/S) + \sigma^2/2 = \ln(K/S) - E^{Q^b_-}[\ln(S_T/S_t)|S_t = S].
\]  

Hence, our first log moneyness measure \( \ell_-(K/S, \sigma \sqrt{\tau}) \) is just the raw difference of the log strike relative from the \( Q^b_- \) mean of the log price relative. Dividing (5) by \( \sigma \sqrt{\tau} \) measures this difference in standard deviations:

\[
z_-(K/S, \sigma \sqrt{\tau}) \equiv \frac{\ell_-(K/S, \sigma \sqrt{\tau})}{\sigma \sqrt{\tau}} = \frac{\ln(K/S) - E^{Q^b_-}[\ln(S_T/S_t)|S_t = S]}{\text{std}^{Q^b_-}[\ln(S_T/S_t)|S_t = S]}.
\]
We think of \( z^- \) as a standardized log moneyness measure under \( Q^b \). We will refer to \( z^- \) as simply moneyness for brevity in the sequel.

To next understand why \( \ell^+ \) is also a log moneyness measure, let \( Q^b^- \) be the equivalent martingale measure arising in the BMS model when one foreign currency unit is instead taken to be the numeraire. Under this probability measure, the log FX rate has positive drift equal to \( \sigma^2/2 \). As a result:

\[
\ell^+(K/S, \sigma \sqrt{\tau}) \equiv \ln(K/S) - \sigma^2 \tau/2 = \ln(K/S) - E^{Q^b^-}[\ln(S_T/S_t)|S_t = S].
\]

Hence, our second log moneyness measure \( \ell^+(K/S, \sigma \sqrt{\tau}) \) is the raw difference of the log strike relative from the \( Q^b^- \) mean of the log price relative. Comparing (5) with (7), the two log moneyness measures are related by:

\[
\ell^+(K/S, \sigma \sqrt{\tau}) = \ell^- (K/S, \sigma \sqrt{\tau}) - \sigma^2 \tau.
\]

Dividing (7) by \( \sigma \sqrt{\tau} \) again measures the difference defining \( \ell^+ \) in standard deviations:

\[
z^+(K/S, \sigma \sqrt{\tau}) \equiv \frac{\ell^+(K/S, \sigma \sqrt{\tau})}{\sigma \sqrt{\tau}} = \frac{\ln(K/S) - E^{Q^b^-}[\ln(S_T/S_t)|S_t = S]}{\text{std}^{Q^b^-}[\ln(S_T/S_t)|S_t = S]} = \sigma \sqrt{\tau}.
\]

Comparing (6) with (9), the two moneyness measures \( z^+ \) and \( z^- \) are related by:

\[
z^+(K/S, \sigma \sqrt{\tau}) = z^- (K/S, \sigma \sqrt{\tau}) - \sigma \sqrt{\tau}.
\]

The log moneyness measures \( \ell^-(K/S, \sigma \sqrt{\tau}) \) and \( \ell^+(K/S, \sigma \sqrt{\tau}) \) are both signed measures. In this paper, we use two other measures of log moneyness which are both non-negative instead. The first non-negative measure of log-moneyness pre-supposes that the two log moneyness measures \( \ell^-(K/S, \sigma \sqrt{\tau}) \) and \( \ell^+(K/S, \sigma \sqrt{\tau}) \) have the same sign, either positive or negative. In this case, the first non-negative measure of log-moneyness arises as their geometric mean:

\[
\bar{\ell}^g(K/S, \sigma \sqrt{\tau}) \equiv \sqrt{\ell^-(K/S, \sigma \sqrt{\tau}) \ell^+(K/S, \sigma \sqrt{\tau})}.
\]

The second non-negative measure of log moneyness pre-supposes in contrast that a call and a put are both OTM using both \( \ell^-(K/S, \sigma \sqrt{\tau}) \) and \( \ell^+(K/S, \sigma \sqrt{\tau}) \). Let \( \ell^c_+ > 0 \) be the OTM call log moneyness and let \( \ell^p_- < 0 \) be the OTM put log moneyness. In this case, our second non-negative measure of log-moneyness is given by:

\[
\bar{\ell}^ag(K/S, \sigma \sqrt{\tau}) \equiv \sqrt{\frac{\ell^c_+ + |\ell^p_-|}{2} \frac{\ell^c_+ + |\ell^p_-|}{2}}.
\]

The subscript \( ag \) on \( \bar{\ell}^ag \) reflects the fact that both arithmetic and geometric means are used.

### 3.3 Relationships between BMS Greeks

In this subsection, we review several well known relationships between Greeks in the zero rates BMS model. The BMS partial differential equation (PDE) is the most well known of these relationships,
relating the first derivative w.r.t. time (henceforth theta) to the second derivative w.r.t. the underlying FX rate (henceforth gamma). When gamma is multiplied by the square of the FX rate, we call the product relative gamma. The zero rates BMS PDE then states that theta is negative one half of the product of relative gamma and the underlying’s instantaneous variance rate.

Other partial derivatives will also be related to relative gamma in this section. When the option premium is differentiated w.r.t. the log of volatility, the result is called relative vega. The product of the variance rate and the second derivative w.r.t. volatility is called relative volga. When relative vega is differentiated w.r.t the log of the underlying FX rate, the result is called relative vanna. In this subsection, we relate relative vega, volga, and vanna to relative gamma.

We will use subscripts on $P^b(S,\sigma,\tau)$ to denote partial derivatives w.r.t. its three arguments. For example, differentiating (1) w.r.t. its first argument $S$ implies that the foreign currency delta of the BMS put is simply:

$$P^b_1(S,\sigma,\tau) = -N(z_+(K/S,\sigma\sqrt{\tau})), \tag{13}$$

since:

$$SN'(z_+(K/S),\sigma\sqrt{\tau}) = KN'(z_-(K/S,\sigma\sqrt{\tau})). \tag{14}$$

Differentiating (13) w.r.t. its first argument $S$ implies that the BMS put gamma is simply:

$$P^b_{11}(S,\sigma,\tau) = \frac{N'(z_+(K/S,\sigma\sqrt{\tau}))}{S\sigma\sqrt{\tau}}. \tag{15}$$

Multiplying (15) by the square of $S$ gives an expression for relative gamma:

$$R\Gamma(S,\sigma,\tau) \equiv S^2P^b_{11}(S,\sigma,\tau) = \frac{SN'(z_+(K/S,\sigma\sqrt{\tau}))}{\sigma\sqrt{\tau}} = \frac{KN'(z_-(K/S,\sigma\sqrt{\tau}))}{\sigma\sqrt{\tau}}, \tag{16}$$

from (14). Notice that relative gamma is measured in the same units as the option premium. All of the relative greeks will have this important property. Hence, when a relative greek can be expressed as a multiple of relative gamma, the factor multiplying relative gamma must be dimensionless. In the rest of this section, we determine these dimensionless multipliers for several greeks of interest.

Differentiating (1) w.r.t. its second argument $\sigma$ implies that the BMS put vega is simply:

$$P^b_2(S,\sigma,\tau) = KN'(z_-(K/S,\sigma\sqrt{\tau}))\sqrt{\tau}, \tag{17}$$

from (14). Multiplying (17) by $\sigma$ gives an expression for the put’s relative vega:

$$R\nu^p(S,\sigma,\tau) \equiv \sigma P^b_2(S,\sigma,\tau) = KN'(z_-(K/S,\sigma\sqrt{\tau}))\sigma\sqrt{\tau}. \tag{18}$$

Multiplying and dividing the RHS of (18) by $\sigma^2\tau$ relates the put’s relative vega to its relative gamma:

$$\sigma P^b_2(S,\sigma,\tau) = R\Gamma(S,\sigma,\tau)\sigma^2\tau, \tag{19}$$

from (16). Thus, for relative vega, the dimensionless multiplier of relative gamma is the total variance $\sigma^2\tau$. 

7
Differentiating (17) w.r.t. its second argument \( \sigma \) implies that the BMS put volga is:

\[
P_{22}^b(S, \sigma, \tau) = \frac{1}{\sigma^2} KN'(z_{-}) z_{-} \ell_{+},
\]

once the arguments of \( z_{-} \) and \( \ell_{+} \) are dropped. Multiplying by \( \sigma^2 \) relates the put’s relative volga to its relative gamma:

\[
\sigma^2 P_{22}^b(S, \sigma, \tau) = KN'(z_{-}) z_{-} \ell_{+} = R \Gamma(S, \sigma, \tau) \ell_{-} \ell_{+},
\]

after dividing and multiplying the middle term of (21) by \( \sigma \sqrt{\tau} \), since \( \ell_{-} = \sigma \sqrt{\tau} z_{-} \). Thus for relative volga, the dimensionless multiplier of relative gamma is the product \( \ell_{-} \ell_{+} = \ell^{2}_{y} \).

Similarly, differentiating (17) w.r.t. its first argument \( S \) implies that the BMS put vanna is:

\[
P_{12}^b(S, \sigma, \tau) = \frac{KN'(z_{-})}{S \sigma} z_{-}.
\]

Recall that put delta is just the negative of the probability of the put finishing in-the-money under \( Q^b_{+} \). As a result, the put delta lies between negative one and zero. When the put is presently OTM, \( z_{-} \) is negative, and hence so is the OTM put’s vanna from (22). Intuitively, a small rise in volatility causes the OTM put’s delta to become more negative, eg. -0.25 becomes -0.26. When the put is presently in-the-money (ITM), \( z_{-} \) is positive, and hence so is the ITM put’s vanna from (22). Intuitively, a small rise in volatility causes the ITM put’s delta to become less negative, eg. -0.75 becomes -0.74. For an ATM (\( \ell_{-} = 0 \)) put, a small rise in volatility has no effect on the put’s delta near negative one half.

Multiplying (22) by \( S \sigma \) relates the put’s relative vanna to its relative gamma:

\[
S \sigma P_{12}^b(S, \sigma, \tau) = \frac{KN'(z_{-})}{\sigma \sqrt{\tau}} z_{-} = R \Gamma(S, \sigma, \tau) \ell_{-}.
\]

After dividing and multiplying the middle term in (23) by \( \sigma \sqrt{\tau} \). Thus for relative vanna, the dimensionless multiplier of relative gamma is simply the log moneyness measure \( \ell_{-} \).

Finally, differentiating (1) w.r.t. its third argument \( \tau \) and negating implies that the BMS put theta is simply:

\[
-P_{3}^b(S, \sigma, \tau) = \frac{-\sigma}{2} \frac{KN'(z_{-})}{\sqrt{\tau}} = \frac{-R \Gamma}{2},
\]

from (16). This last greek relation is known as the BMS PDE.

From put call parity (3), the value \( C_{b}^b(S, \sigma, \tau) \) of a call with the same strike rate and maturity date as the put also satisfies the BMS PDE:

\[
-C_{3}^b(S, \sigma, \tau) = \frac{-R \Gamma}{2} = -P_{3}^b(S, \sigma, \tau).
\]

Also from put call parity (3), the value \( C_{b}^b(S, \sigma, \tau) \) of a call with the same strike rate and maturity date as the put satisfies the same relative greek relations:

\[
\begin{align*}
\sigma C_{2}^b(S, \sigma, \tau) &= R \Gamma(S, \sigma, \tau) \sigma^2 \tau = \sigma P_{2}^b(S, \sigma, \tau) \\
\sigma^2 C_{22}^b(S, \sigma, \tau) &= R \Gamma(S, \sigma, \tau) \ell_{-} \ell_{+} = \sigma^2 P_{22}^b(S, \sigma, \tau) \\
S \sigma C_{12}^b(S, \sigma, \tau) &= R \Gamma(S, \sigma, \tau) \ell_{-} = S \sigma P_{12}^b(S, \sigma, \tau).
\end{align*}
\]
Differentiating put call parity (3) w.r.t. $S$, the foreign currency delta of the call is:

$$C^b_t(S, \sigma, \tau) = 1 + P^b_t(S, \sigma, \tau)$$

$$= 1 - N(z_+(K/S, \sigma \sqrt{\tau})) \text{ from (13)}$$

$$= N(-z_+(K/S, \sigma \sqrt{\tau})).$$

(27)

By the linearity of the greek operators, the straddle value also satisfies the BMS PDE and the same relative greek relations as the put and call. In the next section, we depart from the BMS model, but nonetheless use the BMS greek relations derived in this section to financially interpret the level, slope, and curvature of BMS IV.

4 Assumptions

4.1 Spot FX Dynamics

Let $S_t$ denote the spot FX rate at time $t \in [0, T]$, expressed in domestic per foreign currency unit. We assume that $S_0$ is known and strictly positive. We assume that neither currency can become worthless and that there is no arbitrage between the two currencies. As a result, there exists two equivalent martingale measures, called $Q_-$ and $Q_+$, corresponding to which of the two currencies is taken as numeraire. We call $Q_-$ the domestic measure and we call $Q_+$ the foreign measure. We assume zero interest rates over $[0, T]$, so that $S$ is a local martingale under the domestic measure $Q_-$. We further assume that $S$ is a continuous local martingale with state space $(0, \infty)$ over the time interval $[0, T]$. As a result, there exists a $Q_-$ standard Brownian motion $W$ and a bounded stochastic process $\sigma$ such that $S$ solves the following stochastic differential equation (SDE):

$$dS_t = \sigma_t S_t dW_t, \quad t \in [0, T].$$

(28)

The stochastic process $\{\sigma_t, t \in [0, T]\}$ is called the instantaneous (lognormal) volatility of the spot FX rate. If the process $\sigma$ were constant over time, then the $S$ process solving the SDE (28) would be driftless geometric Brownian motion. However in this paper, $\{\sigma_t, t \in [0, T]\}$ is an arbitrary stochastic process with state space $(0, \infty)$ whose $Q_-$ dynamics are not specified. The only restriction we place on the level and $Q_-$ dynamics of the strictly positive $\sigma$ process is that the $\sigma$ process is bounded, so that one obtains a strictly positive spot FX process $S$. In contrast to classical SV models such as SABR, we do not even assume that $\sigma_0$ is constant. Rather we treat $\sigma_0$ as a positive random variable whose $Q_-$ law is unknown. Thus, the $Q_-$ dynamics for the spot FX rate $S$ described by (28) can be described as driftless geometric Brownian motion, generalized to have arbitrary unspecified volatility.

In SV models, it is common to assume some particular stochastic process for $\sigma$. For example, the Stochastic Alpha Beta Rho (SABR) model of Hagan et al [3] assumes that under $Q_-$, $\sigma$ is a driftless geometric Brownian motion. Any particular specification for the $\sigma$ process leads to a particular specification for the option price process, and hence for the process followed by the IV of the option. In fact, specifying the $Q_-$ dynamics of $\sigma$ over an infinite time horizon leads to a corresponding particular specification for the $Q_-$ dynamics of all IV’s across all strike rates.
$K > 0$ and across all maturity dates $T > 0$. For any fixed maturity date $T$, one output of a stochastic volatility model is therefore a particular specification for the $Q$-dynamics of the IV curve corresponding to that maturity date. The resulting IV dynamics have the virtue of being arbitrage-free, but are never explicit (unless $\sigma$ is constant).

Several groups of researchers have sought to reverse this thought process. In contrast to classical SV models, in the so-called market-modeling approach, one does not assume that investors are able to directly specify the $Q$-dynamics of the $\sigma$ process. Rather, in the market-modeling approach, one directly specifies a particular stochastic process under $Q$ for the IV curve that corresponds to some fixed maturity date. One can alternatively specify the $Q$-dynamics for a single reference IV, for a different curve, eg. the ATM term structure of IV’s, or for the entire IV surface (or swaption cube). Simple and financially natural specification of IV dynamics can then be combined with the implications of no dynamic arbitrage to explicitly determine the dependence of IV on $K$ and/or $T$.

In any of these market-modeling approaches, two difficult problems arise. The first problem is to guarantee that the assumed $Q$-dynamics for IV are free of all arbitrages. Merton’s simple no arbitrage relations for option prices across $K$ and $T$ translate into complicated no arbitrage relations for IV’s. While some researchers are able to specify dynamics which are free of some arbitrages, to the authors’ knowledge, no one has yet succeeded in directly specifying non-trivial IV dynamics which rule out all arbitrages.

Once the first difficult problem of guaranteeing no arbitrage is solved, a second problem arises in the market-modeling approach. This second problem is to determine the implications of no arbitrage for the $Q$-dynamics of the instantaneous volatility process $\sigma$. Since $\sigma$ corresponds to the ATM IV as $T \downarrow t$, this second problem is easiest when the ATM term structure or IV surface is $a$ priori specified. This second problem becomes much harder when the $Q$-dynamics are $a$ priori specified for a strike structure or single IV at some fixed maturity date $T > t$. In particular, when a careless specification of the $Q$-dynamics of a strike structure of IV’s is combined with a no dynamic arbitrage relation, the corresponding implied dynamics of $\sigma$ can depend on $K$. Of course, this result makes no financial sense whenever two or more options trade.

In the next subsection, we partially follow the market-modeling approach by imposing restrictions on the $Q$-dynamics of a strike structure of IV’s at some fixed maturity date $T > t$. However, we bypass the two problems described above by deviating in two different ways. First, we will only partially specify the stochastic process under $Q$ for the IV curve that corresponds to some fixed maturity date. Second, we will not assume no arbitrage across options of different strikes or between options and the two underlying currencies. Put another way, we allow arbitrage whenever a portfolio involves one option or two or more options with different strikes. We assume there is no arbitrage between a put and a call at the same strike rate, i.e put call parity holds at each strike rate $K > 0$. We will also continue to assume that there is no arbitrage between the two currencies in order to give meaning to the equivalent martingale measure $Q$. Since we allow arbitrage whenever an option is involved, the market price of an option will not be obtained by taking the conditional mean under $Q$ of the option’s payoff. Instead, the market price of an option struck at $K$ will be obtained by evaluating the BMS option pricing formula on the IV corresponding to $K$.

Our partial specification of the $Q$-dynamics of a strike structure of IV’s leads to greater generality than the standard market-modeling approach. We are able to work in this greater generality.
because the objective of our paper is not to determine a unique arbitrage-free price of an option. Instead, our objective is to determine how to trade options in a profitable way, whether or not an options market-maker is inadvertently quoting arbitrages via the IV quote generation process described in the next subsection.

4.2 IV Curve Dynamics

To compensate for the absence of a specification of the initial level and dynamics of the instantaneous volatility process $\sigma$, this subsection now directly models the dynamics of IV by strike rate. For some fixed maturity date $T > t$, let $I_t(K) > 0$ be the IV quoted at time $t \geq 0$ for strike rate $K > 0$. By the definition of IV, the corresponding put price $P_t(K)$ is obtained by:

$$P_t(K) = P^b(S_t, I_t(K), T - t), \quad t \in [0, T],$$

where the BMS put pricing function $P^b$ is defined in (1). In contrast to most of the SV literature, we do not assume that the resulting put price process is arbitrage-free. We merely assume that there is no arbitrage between the two currencies so that an equivalent martingale measure $Q_-$ exists. Suppose that under $Q_-$, the stochastic process governing the IV curve $I_t(K)$ at time $t$ is obtained by solving the following SDE:

$$dI_t(K) = \omega_t I_t(K) dZ_t, \quad t \in [0, T), K > 0,$$

where $Z$ is a $Q_-$ standard Brownian motion. Hence, the IV process is assumed to have the same type of dynamics as $S$, i.e continuous and driftless under $Q_-$. We refer to $\omega_t$ as the volvol process. Since the volvol process $\omega$ does not depend on $K$, $\omega_t^2$ is the instantaneous variance rate of the entire IV curve. While the initial IV level $I_0(K)$ need not be flat in $K$, the increments of $\ln I_t(K)$ are flat in $K$ for each $t \in [0, T]$. As a result, the ratio of two IV’s at any future time is just the ratio of the two initial IV’s:

$$\frac{I_t(K)}{I_t(K')} = \frac{I_0(K)}{I_0(K')}, \quad K \neq K', t \in (0, T].$$

We assume that the volvol process is strictly positive, i.e. $\omega_t > 0$. Had we allowed $\omega_t \geq 0$, then at a time $s$ when $\omega_s = 0$, the IV’s are moving by a trivial relative shift. Since $\omega_t > 0$, IV’s by strike are always undergoing non-trivial relative shifts.

If the volvol process $\omega$ were constant over time, then the process for $I(K)$ solving the SDE (30) would be driftless geometric Brownian motion. However, in this paper, $\{\omega_t, t \in [0, T]\}$ is actually an arbitrary stochastic process with state space $(0, \infty)$ whose $Q_-$ dynamics are not specified. In contrast to either classical SV models or market models, we do not even assume that $\omega_0$ is constant. Rather, we treat $\omega_0$ as a strictly positive random variable whose $Q_-$ law is unknown. The only restriction we place on the level and $Q_-$ dynamics of the strictly positive $\omega$ process is that $\omega$ is bounded, so that one obtains a strictly positive IV process at each strike rate $K > 0$. Thus, (28) and (30) imply that the IV curve $I(K)$ has the same type of $Q_-$ dynamics as the spot FX rate, $S$ namely driftless geometric Brownian motion, generalized to have arbitrary unspecified volatility.
4.3 Correlation Process

Recall that $W_t$ and $Z_t$ are the levels at time $t$ of the standard Brownian motions $W$ and $Z$ driving $S_t$ and $I_t(K)$ respectively. For each fixed $t \in [0, T]$, let $\rho_t \in [-1, 1]$ be the bounded random variable governing the instantaneous correlation between the two standard Brownian motions $W$ and $Z$:

$$d\langle W, Z \rangle_t = \rho_t dt, \quad t \in [0, T].$$

(32)

In this paper, $\{\rho_t, t \in [0, T]\}$ is an arbitrary stochastic process with state space $[-1, 1]$ whose $\mathbb{Q}$-dynamics are not specified. In contrast to standard SV models, we do not even assume that $\rho_0$ is some constant in $[-1, 1]$. Rather, we assume that $\rho_0$ is a random variable whose $\mathbb{Q}$-law has support in $[-1, 1]$, but whose $\mathbb{Q}$-law is unknown.

To summarize the results of this section, we assume no knowledge of the probability laws governing the random initial level or the subsequent random changes in the stochastic processes $\sigma$, $\rho$, and $\omega$. We merely require that the initial levels and dynamics of these three stochastic processes do not depend on the strike rate $K > 0$. For $\sigma$ and $\rho$, invariance to $K$ is financially natural, while for $\omega$, this invariance implies that all IV’s change by the same percentage amount. We explore the trading implications of equal percentage shifts because it is the simplest way to parametrize a market-maker’s update process, while keeping IV’s positive.

We will show that when an options market-maker updates his IV’s in the manner indicated by (30), an investor will be able to form three special dynamic option portfolios which each profit on average from some function of the three unspecified stochastic processes $\sigma$, $\rho$, and $\omega$. In particular, the first special option portfolio dynamically trades only ATM straddles. It profits on average from the difference between the instantaneous variance rate $\sigma_t^2$ of log FX and the observed ATM implied variance rate. The second special option portfolio dynamically trades only an equally OTM call and put. It profits on average from the difference between the instantaneous covariation rate $\gamma_t \equiv \sigma_t \rho_t \omega_t$ between the log FX rate and the log IV’s, and the observed slope between two equally OTM halved implied variance rates. Finally, the third special option portfolio dynamically trades only an ATM straddle and an equally OTM call and put. It profits on average from the difference between the instantaneous variance rate $\omega_t^2$ of the log IV’s and an observed convexity measure of the three implied variance rates.

5 Decomposing Instantaneous Gains

We assume that an investor is able to dynamically trade the two currencies. In theory, an investor could also dynamically trade options along the entire continuum of strike rates. To achieve our objectives, a trader will only need to hold options with at most three strike rates at any one time. We assume that at any one time, the trader will in general investor have some position in an out-of-the-money (OTM) put, some position in an ATM straddle, and some position in an OTM call. The investor is allowed to dynamically trade options, so as to vary the strike rates of the three options

---

1When options are priced to be free of arbitrage, the IV’s are determined by $\sigma$, $\rho$, and $\omega$. Since the IV curve has been assumed to have continuous sample paths, this suggests that these three processes cannot jump.
held over time. We assume zero transactions costs. By limiting our trader to holding options with at most three strike rates, we shed light on the discrete strike rate quotation conventions used in OTC options markets for FX and precious metals.

In this section, we first derive expressions for the instantaneous gain on each of the three types of options that can be held at any one time. We then derive an expression for the instantaneous gain on the dynamic portfolio which simultaneously contains options for up to three strike rates.

5.1 Definition of Instantaneous Gain on a Single Option Position

We will refer to a straddle as a single option even though it contains both a call and a put. In this paper, a straddle of maturity date \( T \) is said to be ATM at the fixed time \( t \) and fixed spot level \( S_t \) if the common strike rate \( K \) of the put and call in the straddle is such that:

\[
\ell_-(K/S_t, I_t(K)\sqrt{T-t}) = 0.
\] (33)

Let \( K^a_t \) be this common strike rate and let \( I_{at} \equiv I_t(K^a_t) \) denote the ATM IV quoted at the varying time \( t \geq 0 \) for the fixed maturity date \( T \geq t \), The ATM strike rate \( K^a \) depends on both \( S_t \) and \( I_{at}\sqrt{T-t} \):

\[
K^a(S_t, I_{at}\sqrt{T-t}) \equiv S_t e^{-0.5(I_{at}\sqrt{T-t})^2}.
\] (34)

Let \( I_{pt} \) denote the OTM put IV at time \( t \in [0, T] \). For some fixed strike rate \( K^p_t < K^a_t \), recall from (29) that this OTM put IV is defined so that:

\[
P_t(K^p_t) = P^b(S_t, I_{pt}, T-t; K^p_t), \quad t \in [0, T],
\] (35)

where the dependence of the BMS model put value \( P^b \) on the put’s strike rate \( K^p_t \in (0, K^a_t) \) has now been made explicit. We define the instantaneous gain \( gP_t(K^p_t) \) on the OTM put by:

\[
gP_t(K^p_t) \equiv \left[ dP^b(S_t, I_t(K), T-t; K) \right]_{K=K^p_t}, \quad t \in [0, T].
\] (36)

In words, the instantaneous gain \( gP_t(K^p_t) \) from holding a put at time \( t \in [0, T] \) is just the total derivative w.r.t. time, except that the variation of the strike rate over time has been suppressed. From Itô’s formula, this gain \( gP_t(K^p_t) \) decomposes as:

\[
gP_t(K^p_t) = P^b_t(S_t, I_{pt}, T-t) dS_t + P^b_2(S_t, I_{pt}, T-t) dI_{pt} + \mathcal{G}_t^p dt, \quad t \in [0, T],
\] (37)

where since \( S_t \) and \( I_{pt} \) are \( \mathbb{Q}_- \) local martingales, \( \mathcal{G}_t^p \) is the mean gain rate under \( \mathbb{Q}_- \) on the OTM put at time \( t \in [0, T] \), given by:

\[
\mathcal{G}_t^p = P^b_{11}(S_t, I_{pt}, T-t) \frac{d\langle S \rangle_t}{2 dt} + P^b_{12}(S_t, I_{pt}, T-t) \frac{d\langle S, I_{pt} \rangle_t}{dt} + P^b_{22}(S_t, I_{pt}, T-t) \frac{d\langle I_{pt} \rangle_t}{2 dt} - P^b_3(S_t, I_{pt}, T-t).
\] (38)

From (28) and (30), the second order variations are:

\[
\frac{d\langle S \rangle_t}{dt} = \sigma^2_t S_t^2, \quad \frac{d\langle S, I_{pt} \rangle_t}{dt} = S_t \sigma_{t} \rho \tilde{\omega}_t I_{pt} \equiv S_t \gamma_t I_{pt}, \quad \frac{d\langle I_{pt} \rangle_t}{dt} = \omega^2_t I_{pt}^2.
\] (39)
Substituting (39) in (38) implies that the $\mathbb{Q}_-$ mean gain rate on the OTM put at time $t \in [0, T]$ is:

$$G_t^p = S_t^2 P_{11}^b(S_t, I_{pt}, T - t)\frac{\sigma_t^2}{2} + S_t I_{pt} P_{12}^b(S_t, I_{pt}, T - t)\gamma_t + I_{pt}^2 P_{22}^b(S_t, I_{pt}, T - t)\frac{\omega_t^2}{2} - P_3^b(S_t, I_{pt}, T - t).$$

(40)

Hence, the mean gain on the OTM put at time $t$ is a linear combination of the put’s relative gamma:

$$R \Gamma_t^p = S_t^2 P_{11}^b,$$

(41)

its relative vanna:

$$S_t I_{pt} P_{12}^b = R \Gamma_t^p \ell_{-t}, \quad \ell_{-t} \equiv \ell_-(K_t^p / S_t \sqrt{T - t}),$$

(42)

from (23), its relative volga:

$$I_{pt}^2 P_{22}^b = R \Gamma_t^p \ell_{+t}, \quad \ell_{+t} \equiv \ell_+(K_t^p / S_t \sqrt{T - t}),$$

(43)

from (21), and its theta:

$$-P_3^b = -R \Gamma_t^p \frac{I_{pt}^2}{2},$$

(44)

from (24). Substituting (41) to (44) in (38), the $\mathbb{Q}_-$ mean gain rate on the OTM put at time $t \in [0, T]$ simplifies to:

$$G_t^p = R \Gamma_t^p \left[ \frac{\sigma_t^2}{2} + \gamma_t \ell_{-t} + \frac{\omega_t^2}{2} \ell_{-t} \ell_{-t} + \frac{I_{pt}^2}{2} \right].$$

(45)

Substituting (28), (30), and (45) in (37) implies that the instantaneous gain on the OTM put at time $t \in [0, T]$ is given by:

$$gP_t(K_t^c) = P_t^b(S_t, I_{pt}, T - t)\sigma_t S_t dW_t + P_2^b(S_t, I_{pt}, T - t)\omega_t I_{pt} dZ_t$$

$$+ R \Gamma_t^p \left[ \frac{\sigma_t^2}{2} + \gamma_t \ell_{-t} + \frac{\omega_t^2}{2} \ell_{-t} \ell_{-t} + \frac{I_{pt}^2}{2} \right] dt.$$  

(46)

We now develop the corresponding expressions for the OTM call and for the ATM straddle. Let $I_{ct}$ denote the OTM call IV at time $t \in [0, T]$. For some fixed strike rate $K_t^c > K_t^a$, this OTM call IV is defined so that:

$$C_t(K_t^c) = C_t^b(S_t, I_{ct}, T - t; K_t^c), \quad t \in [0, T],$$

(47)

where the BMS model call pricing function $C_t^b$ is given in (4). The instantaneous gain on the OTM call at time $t \in [0, T]$ is analogously given by:

$$gC_t(K_t^c) = C_t^b(S_t, I_{ct}, T - t)\sigma_t S_t dW_t + C_2^b(S_t, I_{ct}, T - t)\omega_t I_{ct} dZ_t$$

$$+ R \Gamma_t^c \left[ \frac{\sigma_t^2}{2} + \gamma_t \ell_{-t} + \frac{\omega_t^2}{2} \ell_{-t} \ell_{-t} + \frac{I_{ct}^2}{2} \right] dt.$$  

(48)

Let $I_{at}$ denote the ATM put IV at time $t \in [0, T]$. This ATM put IV is defined so that:

$$P_t(K_t^a) = P_t^b(S_t, I_{at}, T - t; K_t^a), \quad t \in [0, T].$$

(49)
Recall we assumed that put-call-parity holds at all strikes $K > 0$. Since put-call-parity holds in particular at the ATM strike rate $K^a$, the ATM call IV and the ATM straddle IV are both given by $I_{at}$. Let $A^b(S, \sigma, \tau; K) \equiv P^b(S, \sigma, \tau; K) + C^b(S, \sigma, \tau; K)$ be the formula for the BMS value of an ATM straddle. For $t \in [0, T]$, let $A_t(K^a) \equiv A^b(S_t, I_{at}, T - t; K^a)$ be the stochastic process describing the price of the ATM straddle maturing at $T$ in our SV setting. The gains on the ATM put and call at time $t \in [0, T]$ are analogous to the gains on the OTM options, except that $\ell_{t,t}^a = 0$. The instantaneous gain on the ATM straddle at time $t \in [0, T]$ is therefore given by the simpler expression:

$$gA_t(K^a_t) = A_t^b(S_t, I_{at}, T - t) \sigma_t S_t dW_t + A_t^b(S_t, I_{at}, T - t) \omega_t I_{at} dZ_t$$

\[+ R\Gamma_t^a \left[ \frac{\sigma_t^2}{2} - \frac{\ell_{at}^2}{2} \right] dt, \quad (50)\]

where $R\Gamma_t^a \equiv S_t^2 A_{11}^b(S_t, I_{at}, T - t) = S_t^2 [P_{11}^b(S_t, I_{at}, T - t) + C_{11}^b(S_t, I_{at}, T - t)]$ is the relative gamma of the ATM straddle. The foreign currency delta of the ATM straddle is given by:

$$\Delta_t^a = P^b_1(S_t, I_{at}, T - t) + C^b_1(S_t, I_{at}, T - t)$$

$$= -N(\ell_{+t}^a) + N(-\ell_{-t}^a) \text{ from (13) and (27)}$$

$$= N(I_{at} \sqrt{T - t}) - N(-I_{at} \sqrt{T - t}), \quad (51)$$

from setting $\ell_{t,t}^a = 0$ and $\sigma = I_{at}$ in (10). Since we defined ATM to zero out the $\ell_-$ moneyness measure rather than the $\ell_+$ moneyness measure, our ATM straddle has non-zero delta with respect to the foreign currency. However, our ATM straddle has zero delta with respect to the so-called domestic currency, so we are in line with the ATM definition in the OTC FX option market.

### 5.2 Instantaneous Gain of a Three Strike Rate Option Portfolio

Let $n^p_t$ be the number of OTM puts held with strike rate $K^p_t < K^a(S_t, I_{at} \sqrt{T - t})$ at time $t \in [0, T]$. Similarly, let $n^a_t$ be the number of ATM straddles held with strike rate $K^a_t = K^a(S_t, I_{at} \sqrt{T - t})$ at time $t \in [0, T]$. Finally, let $n^c_t$ be the number of OTM calls held with strike rate $K^c_t > K^a(S_t, I_{at} \sqrt{T - t})$ at time $t \in [0, T]$. Let $V_t$ be the value of the three strike rate option portfolio at time $t \in [0, T]$:

$$V_t \equiv n^p_t P_t(K^p_t) + n^a_t A_t(K^a_t) + n^c_t C_t(K^c_t), \quad t \in [0, T]. \quad (52)$$

The instantaneous gain on this portfolio at time $t \in [0, T]$ is defined as:

$$gV_t \equiv n^p_t gP_t(K^p_t) + n^a_t gA_t(K^a_t) + n^c_t gC_t(K^c_t), \quad t \in [0, T]. \quad (53)$$

Thus, the instantaneous gain $gV_t$ differs from the total derivative $dV_t$ in that the former suppresses the variation of the holdings over time. As a result, the gain in value of the option portfolio is just a linear combination of the gains in each option price.
Substituting (46), (48), and (50) in (53) implies that the instantaneous gain on the option portfolio at time \( t \in [0, T] \) is given by:

\[
gV_t = \mathcal{G}_t^v dt + \Delta_t^v \sigma_t dW_t + Rv_t^v \omega_t dZ_t, \tag{54}
\]

where \( \mathcal{G}_t^v \) is the \( \mathbb{Q}_- \) mean gain rate on the option portfolio at time \( t \in [0, T] \):

\[
\mathcal{G}_t^v = \eta_t^p R \Gamma_t^p \left[ \frac{\sigma_t^2}{2} + \sqrt{T - t} \gamma_t \ell_{-t}^p + (T - t) \frac{\omega_t^2}{2} \ell_{-t}^p \ell_{t}^p - \frac{I_{pt}^2}{2} \right] \\
+ \eta_t^a R \Gamma_t^a \left[ \frac{\sigma_t^2}{2} - \frac{I_{at}^2}{2} \right] \\
+ \eta_t^c R \Gamma_t^c \left[ \frac{\sigma_t^2}{2} + \sqrt{T - t} \gamma_t \ell_{-t}^c + (T - t) \frac{\omega_t^2}{2} \ell_{-t}^c \ell_{t}^c - \frac{I_{ct}^2}{2} \right], \tag{55}
\]

\( \Delta_t^v \) is the foreign currency delta of the option portfolio at time \( t \in [0, T] \):

\[
\Delta_t^v = \eta_t^b P_t^b(S_t, I_{pt}, T - t) + \eta_t^a A_t^b(S_t, I_{at}, T - t) + \eta_t^c C_t^b(S_t, I_{ct}, T - t) \\
= -\eta_t^b N(\ell_{-t}^b) + \eta_t^a \left[ N(I_{at}\sqrt{T - t}) - N(-I_{at}\sqrt{T - t}) \right] + \eta_t^c N(-\ell_{-t}^c), \tag{56}
\]

from (13), (51), and (27), while \( Rv_t^v \) is the relative vega of the option portfolio at time \( t \in [0, T] \):

\[
Rv_t^v = \eta_t^b P_t^b(S_t, I_{pt}, T - t) + \eta_t^a A_t^b(S_t, I_{at}, T - t) + \eta_t^c C_t^b(S_t, I_{ct}, T - t) \\
= (T - t) \left[ \eta_t^b R \Gamma_t^p I_{pt}^2 + \eta_t^a R \Gamma_t^a I_{at}^2 + \eta_t^c R \Gamma_t^c I_{ct}^2 \right]. \tag{57}
\]

using (19). The vega \( v_t^v \) of the option portfolio is given by:

\[
v_t^v = \eta_t^b P_t^b(S_t, I_{pt}, T - t) + \eta_t^a A_t^b(S_t, I_{at}, T - t) + \eta_t^c C_t^b(S_t, I_{ct}, T - t) \\
= (T - t) \left[ \eta_t^b R \Gamma_t^p I_{pt} + \eta_t^a R \Gamma_t^a I_{at} + \eta_t^c R \Gamma_t^c I_{ct} \right], \tag{58}
\]

by dividing (19) by \( \sigma > 0 \). Note from (57) and (58) that the relationship between relative vega and relative gamma is similar to the relationship between vega and relative gamma. However, the former relationship involves implied variance rates, while the latter relationship involves IV’s. This distinction will become relevant when we motivate OTC FX option quotations which involve IV rather than implied variance rates.

Equation (55) indicates that in general, the \( \mathbb{Q}_- \) mean gain rate on an option portfolio at time \( t \in [0, T] \) depends on the instantaneous variance rate \( \sigma _t^2 \) of the underlying log FX rate, the covariation rate \( \gamma _t \equiv \sigma _t \rho _t \omega _t \) between the log FX rate and the log IV’s, and the instantaneous variance rate \( \omega _t^2 \) of the log IV’s. Recall our assumption that each of these three stochastic processes is both unspecified and unobserved ex ante. In the next section, we will show that despite this unobservability, a trader can create three special dynamic option portfolios whose \( \mathbb{Q}_- \) mean gain rate depends on only one of these stochastic processes.
6 Vol, Skew, and Smile Trades

In this section, we define three special dynamic option portfolios called the vol, skew, and smile trade. For each trade, we calculate the \( Q \)-mean gain rate and the vega. We find that for each trade, the \( Q \)-mean gain rate only depends on one of the unobservable processes \( \sigma_t^2, \gamma_t \equiv \sigma_t \rho \omega_t, \) and \( \omega_t^2 \). We also find that the vega of each trade is positively proportional to one of the quotes called ATM straddle, risk-reversal, and butterfly-spread.

6.1 Vol Trade

In this subsection, we seek a dynamic option portfolio whose \( Q \)-mean gain rate at time \( t \in [0, T] \) depends on the instantaneous variance rate, \( \sigma_t^2 \), of the log FX rate, but is not exposed to either the instantaneous covariation rate \( \gamma_t \), of the log FX rate with each log IV, or to the instantaneous variance rate \( \omega_t^2 \) of each log IV.

Consider the following option positions:

\[
\begin{align*}
\eta^p_t &= 0 \\
\eta^a_t &= 2 \\
\eta^c_t &= 0,
\end{align*}
\tag{59}
\]

We refer to this position as long \( \frac{2}{\text{ATM}} \) units of an ATM straddle. Substituting (59) in (55) implies that the \( Q \)-mean gain rate of this position at time \( t \in [0, T] \) is:

\[
G^v_t = \sigma_t^2 - I^2_{at}, \quad t \in [0, T].
\tag{60}
\]

Since the RHS of (60) does not depend on \( \gamma_t \) or \( \omega_t^2 \), we have found a dynamic portfolio whose \( Q \)-mean gain rate at time \( t \in [0, T] \) depends only on \( \sigma_t^2 \). The quantity \( \sigma_t^2 \) received in (60) is the instantaneous variance rate of \( \ln S \). The quantity \( I^2_{at} \) subtracted in (60) is the BMS ATM implied variance rate. Equation (60) helps clarify why the prices of ATM options with V shaped payoffs are thought to reflect the variance of future values of its underlying, rather than the mean absolute deviation of these future values.

Substituting (59) in (58) implies that the vega of the ATM straddle position is positively proportional to the ATM IV:

\[
v^v_t = 2(T - t)I_{at}, \quad t \in [0, T].
\tag{61}
\]

Recall that we have defined ATM to mean \( \ell_{-t} = 0 \). In OTC FX option markets, ATM is associated with zero straddle delta, which corresponds to \( \ell_{+t} = 0 \) instead. However, if we define \( \hat{S} = \frac{1}{S} \) and \( \hat{K} = \frac{1}{K} \), then \( \ell_{-t} = 0 \) is equivalent to \( \ell_{+t} = 0 \), matching the quoting convention in FX options markets.

Equation (60) for the \( Q \)-mean gain rate justifies referring to the position (59) as a trade on the variance rate of log FX. However, (61) for the vega justifies referring to it as a vol trade instead. We have chosen the shorter name in this paper. From the chain rule, the first derivative of the position value w.r.t. the variance rate \( \sigma^2 \) is simply the time to maturity. It is difficult to imagine a cleaner result.
6.2 Skew Trade

In this subsection, we seek a dynamic option portfolio whose $Q_-$ mean gain rate at time $t \in [0, T]$ depends on the instantaneous covariation rate, $\gamma_t \equiv \sigma_t \rho_t \omega_t$, of the log FX rate with each log IV, but is not exposed to either the instantaneous variance rate, $\sigma_t^2$, of the log FX rate or to the instantaneous variance rate, $\omega_t^2$, of each log IV.

Suppose first that the OTM call strike rate $K_c^t$ is such that $\ell_{\pm t}^c > 0$, while the OTM put strike rate $K_p^t$ is such that $\ell_{\pm t}^p < 0$. We now suppose that the put and the call are equally OTM using geometric mean log moneyness:

$$\sqrt{\ell_{-t}^p \ell_{+t}^p} = \sqrt{\bar{\ell}_{-t} \ell_{+t}^c} \equiv \bar{\ell}_{gt}, \quad t \in [0, T],$$

from (11). It is straightforward to show that for any OTM call strike $K_c^t$ and corresponding IV $I_{ct}$ leading to particular numerical values for $\ell_{-t}^c$ and $\ell_{+t}^c$, the OTM put strike solving (62) is explicitly given by:

$$K_{pt} = S_t e^{-\sqrt{2(\ell_{-t}^c \ell_{+t}^c)} + I_{pt} \tau^2/2} \quad t \in [0, T].$$

(63)

Under the strike restrictions, $\ell_{\pm t}^c > 0$, $\ell_{\pm t}^p < 0$, and (62), (55) implies that the $Q_-$ mean gain rate on the option portfolio at time $t \in [0, T]$ is given by:

$$G_v^t = \eta^p_t R \Gamma_p^t \left[ \frac{\sigma_t^2}{2} + \gamma_t \ell_{-t}^p + \frac{\omega_t^2}{2} \ell_{gt}^2 - \frac{I_{pt}^2}{2} \right]$$

$$+ \eta^a_t R \Gamma_a^t \left[ \frac{\sigma_t^2}{2} - \frac{I_{at}^2}{2} \right]$$

$$+ \eta^c_t R \Gamma_c^t \left[ \frac{\sigma_t^2}{2} + \gamma_t \ell_{-t}^c + \frac{\omega_t^2}{2} \ell_{gt}^2 - \frac{I_{ct}^2}{2} \right].$$

(64)

Consider the following option positions:

$$\eta^p_t = -\frac{1}{(\ell_{-t}^c - \ell_{+t}^p) R \Gamma_p^t} \quad \eta^a_t = 0 \quad \eta^c_t = \frac{1}{(\ell_{-t}^c - \ell_{+t}^p) R \Gamma_c^t} \quad t \in [0, T].$$

(65)

Since $\ell_{-t}^c - \ell_{-t}^p > 0$, this is a long position of $\frac{1}{\ell_{-t}^c - \ell_{-t}^p}$ normalized risk-reversals, each with value $\frac{C_t(K_c^t) - P_t(K_p^t)}{R \Gamma_t^c}$ at time $t \in [0, T]$. Recall that the ratio of OTM calls to OTM puts in a standard risk-reversal is minus one. The long position $\frac{1}{\ell_{-t}^c - \ell_{-t}^p} > 0$ is said to be in normalized risk-reversals because each option position is first normalised to have unit relative gamma and then the standard risk-reversal ratio is applied.

Substituting (65) in (64) implies that the $Q_-$ mean gain rate at time $t \in [0, T]$ of the long position in normalized risk-reversals is:

$$G_v^t = \gamma_t - b_t,$$

(66)

where:

$$b_t \equiv \frac{I_{pt}^2}{2} - \frac{I_{pt}^2}{2} \frac{\ell_{-t}^p}{\ell_{-t}^c - \ell_{-t}^p}$$

(67)

18
is a finite difference of halved implied variance rates by the log moneyness measure $\ell_{-t}$. Clearly, $b_t$ is just the slope of the line connecting the two equally OTM halved implied variance rates $\frac{I_{ct}^2}{2}$ and $\frac{I_{pt}^2}{2}$, when all halved implied variance rates are plotted against the log moneyness measure $\ell_{-t}$.

From (66), the quantity $\gamma_t$ received in return for paying $b_t$ is the covariation rate between the log FX rate and each log IV. Substituting (65) in (58) implies that the vega of the long position in normalized risk-reversals is just the product of the term $T - t$ and the slope of IV by log moneyness $\ell_{-t}$:

$$v^v_t = (T - t) \frac{I_{ct} - I_{pt}}{\ell^c_{-t} - \ell^p_{-t}} \quad t \in [0, T].$$

Practitioners refer to any measure of the IV slope as the skew. Since $v^v_t = \frac{T - t}{\ell^c_{-t} - \ell^p_{-t}} (I_{ct} - I_{pt})$, the vega is also positively proportional to the risk-reversal quote $I_{ct} - I_{pt}$. Note that the standard risky quote uses IV by delta rather than IV by log moneyness. However, the standard risky quote is a difference of IV’s, rather than a slope of IV’s. As a result, $I_{ct} - I_{pt}$ can also be interpreted as a difference of IV’s by delta, making it identical to the standard FX risky quote.

Equation (66) for the $Q_-$ mean gain rate justifies calling the long position in normalized risk-reversals a covariation rate trade. However, (68) for the vega justifies referring to it as a skew trade instead. We have chosen the shorter name in this paper.

### 6.3 Smile Trade

In this subsection, we seek a dynamic option portfolio whose $Q_-$ mean gain rate at time $t \in [0, T]$ depends on the instantaneous variance rate of log IV, $\omega_t^2$, but is not exposed to either the instantaneous variance rate of log FX, $\sigma_t^2$, or to the instantaneous covariation rate of log FX with each log IV, $\gamma_t$.

Suppose we restrict the OTM put and call strike rates $K^p_t$ and $K^c_t$ so that call log moneyness is positive under both measures, $\ell^c_{-t} > 0$, while put log moneyness is negative under both measures, $\ell^p_{-t} < 0$. Recall from (12) the following non-negative measure of log moneyness:

$$\bar{\ell}_{agt} \equiv \sqrt{\frac{\ell^c_{-t} + |\ell^p_{-t}|}{2}} \frac{\ell^c_{+t} + |\ell^p_{+t}|}{2} > 0 \quad t \in [0, T].$$

Now suppose we restrict the OTM put and call strike rates $K^p_t$ and $K^c_t$ so that the call and put are equally OTM using the $\ell_-$ measure of log moneyness:

$$\ell^c_{-t} = -\ell^p_{-t} \equiv \ell_{-t} > 0 \quad t \in [0, T].$$

It is straightforward to show that for any OTM call strike $K^c_t$ and corresponding IV $I_{ct}$ leading to a particular numerical value for $\ell^c_{-t}$, the OTM put strike solving (70) is explicitly given by:

$$K_{pt} = S_t e^{-\sqrt{2(\ell^c_{-t})^2 + I_{pt}^2/2}} \quad t \in [0, T].$$
Substituting the equal moneyness restriction (70) in (69), the log moneyness measure \( \ell_{agt} \) simplifies to:

\[
\ell_{agt} = \sqrt{\ell_{-t} - \ell_{+t}^2} > 0 \quad t \in [0, T].
\] (72)

Substituting the equal moneyness restriction (70) in (55), the \( Q_- \) mean gain rate of the option portfolio at time \( t \in [0, T] \) simplifies slightly to:

\[
G'_t = \eta^p_t R\Gamma^p_t \left[ \frac{\sigma^2}{2} - \gamma_t \ell_{-t} - \frac{\omega^2}{2} \ell_{-t} \ell_{+t}^p - \frac{I^2_{pt}}{2} \right] + \eta^a R\Gamma^a_t \left[ \frac{\sigma^2}{2} - \frac{I^2_{at}}{2} \right] + \eta^c R\Gamma^c_t \left[ \frac{\sigma^2}{2} + \gamma_t \ell_{-t} + \frac{\omega^2}{2} \ell_{-t} \ell_{+t}^c - \frac{I^2_{ct}}{2} \right]
\] (73)

Consider the following option positions:

\[
\begin{align*}
\eta^p_t &= \frac{1}{\ell_{agt}^2 R\Gamma^p_t} \\
\eta^a_t &= -\frac{2}{\ell_{agt}^2 R\Gamma^a_t} \\
\eta^c_t &= \frac{1}{\ell_{agt}^2 R\Gamma^c_t}
\end{align*}
\] (74)

The portfolio holdings in (74) describe a long position of \( \frac{1}{\ell_{agt}} \) normalized butterfly-spreads, each with value \( \frac{C_t(K^c_t)}{R\Gamma^c_t} - 2\frac{A_t(K^p_t)}{R\Gamma^p_t} + \frac{P_t(K^p_t)}{R\Gamma^p_t} \). Recall that a standard butterfly-spread combines a long position in a strangle with a short position in 2 ATM straddles. The long position \( \frac{1}{\ell_{agt}} > 0 \) is said to be in normalized butterfly-spreads because each option position is first normalised to have unit relative gamma, and then the standard butterfly-spread ratios are applied.

Substituting (74) in (73) implies that the \( Q_- \) mean gain rate at time \( t \in [0, T] \) of the long position in normalized butterfly-spreads is:

\[
G'_t = \omega^2_t - c_t,
\] (75)

where:

\[
c_t \equiv \frac{\ell_{agt}^2 R\Gamma_t^a - I^2_{at}}{I^2_{agt}^2}
\] (76)

is half the centered second order finite difference of implied variance by RMS moneyness. Clearly, \( c_t \) is a measure of the convexity of halved implied variance rates:

\[
\frac{\ell_{agt}^2 R\Gamma_t^a - I^2_{at}}{\ell_{agt}^2} = \frac{\ell_{agt}^2 R\Gamma_t^a - I^2_{at}}{\ell_{agt}^2}
\] (77)

From (75), the quantity \( \omega^2_t \) received in return for paying \( c_t \) is the variance rate of each log IV.
Substituting (74) in (58) implies that the vega of the long position in normalized butterfly-spreads is just half the product of the term $T - t$ and the second order finite difference, $\frac{I_{ct} - 2I_{at} + I_{pt}}{\ell_{agt}^2}$:

$$v_t^v = (T - t) \frac{I_{ct} + I_{pt} - I_{at}}{\ell_{agt}^2} = \frac{T - t}{2} \frac{I_{ct} - 2I_{at} + I_{pt}}{\ell_{agt}^2} \quad t \in [0, T].$$

(78)

The quantity $\frac{I_{ct} - 2I_{at} + I_{pt}}{\ell_{agt}^2}$ is called a standard second order finite difference because it is a finite difference of finite differences:

$$\frac{I_{ct} - 2I_{at} + I_{pt}}{\ell_{agt}^2} = \frac{I_{ct} - I_{at}}{\ell_{agt}} - \frac{I_{at} - I_{pt}}{\ell_{agt}} \quad t \in [0, T].$$

(79)

Hence, (78) implies that the vega of the long position in normalized butterfly-spreads is positively proportional to a convexity measure of IV by moneyness. Practitioners refer to any measure of convexity of the IV curve as the smile. Since $v_t^v = \frac{T - t}{\ell_{agt}^2} (\frac{I_{ct} + I_{pt}}{2} - I_{at})$ from (78), the vega is also positively proportional to the standard FX fly quote $\frac{I_{ct} + I_{pt}}{2} - I_{at}$, which is just the difference between the average OTM IV and the ATM IV. Note that the standard fly quote uses IV by delta rather than IV by log moneyness. However, the standard fly quote is a difference of differences, rather than a slope of slopes. As a result, $\frac{I_{ct} + I_{pt}}{2} - I_{at}$ can also be interpreted as a difference of differences of IV by delta, making it identical to the standard FX fly quote.

Equation (75) for the $Q_-$ mean gain rate justifies calling the long position in normalized butterfly-spreads a variance rate of log IV trade. However, (78) for the vega justifies referring to it as a smile trade instead. We have chosen the shorter name in this paper.

7 Statistical Arbitrage

In the last section, we developed three option portfolios called the vol, skew, and smile trade respectively. For each of these three trades, this section describes a simple sufficient condition on the level, slope, or curvature of the three IV quotes that leads to positive drift under $Q_-$ in the value of the vol, skew, or smile trade respectively. Since the three trades are in general still risky, we refer to this type of trade as a statistical arbitrage. In the next section, we will explore riskless arbitrage for just skew trading and smile trading. For both of these trading strategies, we develop a more complicated sufficient condition and a more complicated option portfolio that generates riskless arbitrage.

7.1 Positive Lower Bound on Instantaneous Volatility

Recall that we treat the instantaneous volatility $\sigma_t$ as a random variable with positive variance, even at time $t = 0$. In this subsection only, we assume that the investor knows at some fixed time $t \in [0, T]$ that the random instantaneous volatility, $\sigma_t$, exceeds the observed ATM IV, $I_{at}$:

$$\sigma_t > I_{at}.$$ 

(80)
Recall from (60) that the vol trade induces the following $\mathbb{Q}$ drift in the value of the position:

$$G_t^v = \sigma_t^2 - I_t^2 \quad t \in [0, T].$$  \hfill (81)

From (80), this $\mathbb{Q}$ drift is a positive random variable. Thus, holding $\frac{1}{2\gamma_t}$ ATM straddles at time $t \in [0, T]$ is a statistical arbitrage when (80) holds.

Substituting (59) in (56) implies that the foreign currency delta of this ATM straddle position is:

$$\Delta_t^v = \frac{2[N(I_{at\sqrt{T}} - t) - N(-I_{at\sqrt{T}} - t)]}{R\Gamma_t^a}, \quad t \in [0, T].$$  \hfill (82)

Since we defined ATM to zero out the $\ell_-$ moneyness measure rather than the $\ell_+$ moneyness measure, our ATM straddle has non-zero delta with respect to the foreign currency. However, recall that the ATM straddle has zero delta with respect to the so-called domestic currency, so our ATM definition agrees with the convention used in the OTC FX option market. Nonetheless, the fact that $\Delta_t^v$ in (82) does not vanish implies that the instantaneous gains on the statistical arbitrage are risky when measured in the domestic currency.

Substituting (59) in (57) implies that the relative vega of the position is:

$$Rv_t^v = (T - t)I_t^2 \quad t \in [0, T].$$  \hfill (83)

Since neither the foreign currency delta nor the relative vega vanish, the normalized ATM straddle is not a riskless arbitrage.

### 7.2 Lower Bound on Instantaneous Covariation Rate

The instantaneous covariation rate between the log FX rate $\ln S$ and the log IV’s $\ln I$ at time $t$ is given by $\gamma_t = \sigma_t \rho_t \omega_t$. Recall that we treat this instantaneous covariation rate as a random variable with positive variance, even at time $t = 0$. Also recall the quantity $b_t$ defined in (67):

$$b_t = \frac{I_{at\sqrt{T}}^2 - I_{at\sqrt{T}}^2}{\ell_{-t}^c - \ell_{-t}^p} \quad t \in [0, T].$$  \hfill (84)

The raw distance between the OTM call log moneyness $\ell_{-t}^c > 0$ and the OTM put log moneyness $\ell_{-t}^p < 0$ is the denominator in (84), $\ell_{-t}^c - \ell_{-t}^p > 0$. Suppose that halved implied variance rates are plotted against the log moneyness measure $\ell_{-t}$. Then $b_t$ in (84) is the slope of the line connecting the two OTM halved implied variance rates.

Suppose that instead of (80), we assume in this subsection that the investor knows that at some fixed time $t \in [0, T]$, this slope is dominated by the random covariation rate $\gamma_t$:

$$\gamma_t > b_t.$$  \hfill (85)

In this case, recall from (66) that the skew trade induces the following $\mathbb{Q}$ drift in the value of the position:

$$G_t^v = \gamma_t - b_t.$$  \hfill (86)
When (85) holds, (86) implies that the $\mathbb{Q}_-$ drift at time $t \in [0, T]$ in the value of the normalized risk-reversal is a positive random variable:

$$\mathcal{G}_t^p > 0.$$  (87)

Thus, the skew trade (65) is a statistical arbitrage when (85) holds.

Substituting (65) in (56) implies that the foreign currency delta of the long position in normalized risk-reversals is:

$$\Delta^p_t = \frac{1}{\ell^p_t - \ell^p_t} \left[ \frac{N(\ell^p_t)}{R_t^p} + \frac{N(-\ell^c_t)}{R_t^c} \right] > 0 \quad t \in [0, T].$$  (88)

This is the slope of the line connecting the two normalized OTM deltas when they are plotted against the log moneyness measure $\ell_{-t}$.

Substituting (65) in (57) implies that the relative vega of the long position in normalized risk-reversals is:

$$Rv_t^p = (T - t) \frac{I^2_{ct} - I^2_{pt}}{\ell^c_t - \ell^p_t} \quad t \in [0, T].$$  (89)

Thus, the relative vega of the skew trade is positively proportional to $\frac{I^2_{ct} - I^2_{pt}}{\ell^c_t - \ell^p_t}$, which is the slope of the line connecting the two OTM implied variances in a plot against the log moneyness measure $\ell_{-t}$.

Substituting (65) in (57) implies that the relative vega of the long position in normalized risk-reversals is:

$$Rv_t^p = (T - t) \frac{I^2_{ct} - I^2_{pt}}{\ell^c_t - \ell^p_t} \quad t \in [0, T].$$  (89)

Thus, the relative vega of the skew trade is positively proportional to $\frac{I^2_{ct} - I^2_{pt}}{\ell^c_t - \ell^p_t}$, which is the slope of the line connecting the two OTM implied variances in a plot against the log moneyness measure $\ell_{-t}$. If $I_{pt} \neq I_{ct}$, then relative vega does not vanish, so the skew trade is not a riskless arbitrage.

If $I_{pt} = I_{ct}$, then $b_t = 0$, so (85) holds when $\gamma_t > 0$, or equivalently when the correlation $\rho_t$ is strictly positive. Thus, the skew trade that generates positive $\mathbb{Q}_-$ drift under positive correlation yet equal OTM IV’s is just a position in $\frac{1}{\ell^c_t - \ell^p_t}$ normalized risk-reversals. Furthermore, if $I_{pt} = I_{ct}$, then the equal geometric moneyness condition (62) implies that the geometric mean of the OTM call strike rate and the OTM put strike rate used in the normalized risk-reversal is just the spot FX rate:

$$\sqrt{K^c_t K^p_t} = S_t \quad t \in [0, T].$$  (90)

From (7), (9), and (16), the relative gammas at time $t \in [0, T]$ of the equally OTM put and call are:

$$R\Gamma^p_t = \frac{S_t}{I_{pt} \sqrt{T - t}} N'\left( \frac{\ln(K^p_t / S)}{I_{pt} \sqrt{T - t}} - \frac{1}{2} I_{pt} \sqrt{T - t} \right), \quad R\Gamma^c_t = \frac{S_t}{I_{ct} \sqrt{T - t}} N'\left( \frac{\ln(S / K^c_t)}{I_{ct} \sqrt{T - t}} + \frac{1}{2} I_{ct} \sqrt{T - t} \right).$$  (91)

If $I_{pt} = I_{ct}$, then substituting (90) in (91) implies that the relative gammas at time $t \in [0, T]$ of the equally OTM put and call are the same:

$$R\Gamma^p_t = R\Gamma^c_t = R\Gamma_t.$$  (92)

Thus, when $I_{pt} = I_{ct}$, the skew trade (65) can be interpreted as a long position of $\frac{1}{(\ell^c_t - \ell^p_t)R_t}$ standard risk-reversals, each with value $C_t(K^c_t) - P_t(K^p_t)$.

Furthermore, if $I_{pt} = I_{ct}$, then (89) implies that the relative vega of this position vanishes. However, one must further delta-hedge the long position in risk-reversals to generate riskless arbitrage. We assume that a forward contract exists on the underlying FX rate. We need to determine
the number, $\eta^f_t$ of forward contracts to short at time $t \in [0, T]$. Note that if $I_{pt} = I_{ct}$, then $\ell^c_t - \ell^p_t = \ell^c_{+t} - \ell^p_{+t}$, so from (88) and (92), the foreign currency delta of the long position in risk-reversals is also:

$$\Delta^c_t = \frac{N(\ell^c_{+t}) + N(-\ell^c_{+t})}{(\ell^c_{+t} - \ell^p_{+t})R\Gamma_t} > 0, \quad t \in [0, T].$$

(93)

This is the slope of the line connecting the two normalized OTM deltas when they are plotted against the log moneyness measure $\ell_{+t}$. As a result, the number of forward contracts to open is given by:

$$\eta^f_t = -\frac{N(\ell^p_{+t}) + N(-\ell^c_{+t})}{(\ell^c_{+t} - \ell^p_{+t})R\Gamma_t} < 0, \quad t \in [0, T].$$

(94)

We conclude that when $\rho_t > 0$ and yet $I_{pt} = I_{ct}$, the delta-hedged skew trade is both a statistical arbitrage and a riskless arbitrage. Obviously if $\rho_t < 0$ and yet $I_{pt} = I_{ct}$, the negative of the delta hedged skew trade is both a statistical arbitrage and a riskless arbitrage. When $I_{pt} = I_{ct}$, temporarily imposing no arbitrage among options at different strike rates implies that $\gamma_t = 0$ so $\rho_t = 0$. Carr and Lee[2] prove the converse of this statement.

### 7.3 Positive Lower Bound on Instantaneous Vol of Vol

Recall that we treat the instantaneous volatility of IV, $\omega_t$, as a random variable with positive variance, even at time $t = 0$. Also recall the quantity $c_t$ defined in (76):

$$c_t \equiv \frac{I^2_{pt} + I^2_{ct}}{2} - I^2_{at}, \quad t \in [0, T].$$

(95)

Recall that $c_t$ measures the convexity of the three halved implied variance rates by the log moneyness measure $\bar{\ell}_{agt}$. Suppose that instead of (85), we assume in this subsection that the investor knows at some fixed time $t \in [0, T]$ that this convexity measure of halved implied variance rates is dominated by the random variance rate of log IV, $\omega^2_t$:

$$\omega^2_t > c_t.$$ 

(96)

In this case, recall from (75) that the smile trade induces the following $\mathbb{Q}_-$ drift in the value of the position:

$$G^v_t \equiv \omega^2_t - c_t, \quad t \in [0, T].$$

(97)

When (96) holds, (97) implies that the $\mathbb{Q}_-$ drift in the value of the option portfolio is a positive random variable:

$$G^v_t > 0, \quad t \in [0, T].$$

(98)

Thus, the smile trade (74) is a statistical arbitrage when (96) holds.

Substituting (74) in (56) implies that the foreign currency delta of the smile trade is:

$$\Delta^v_t = \frac{1}{\ell^2_{agt}} \left\{ \frac{N(-\ell^c_{+t})}{R\Gamma^c_t} - 2\frac{N(I_{at}\sqrt{T-t}) - N(I_{at}\sqrt{T-t})}{R\Gamma^a_t} - \frac{N(\ell^p_{+t})}{R\Gamma^p_t} \right\}, \quad t \in [0, T].$$

(99)
This is a convexity measure of the three normalized deltas when they are plotted against the log moneyness measure $\bar{\ell}_{agt}$.

Substituting (74) in (57) that the relative vega of the smile trade is:

$$Rv^v_t = (T-t) \frac{I^2_{ct} + I^2_{pt}}{2 \bar{\ell}^2_{agt}} - I^2_{at}, \quad t \in [0, T].$$

(100)

The centered second order finite difference of implied variance rates by log moneyness $\bar{\ell}_{agt}$ is given by $\frac{I^2_{ct} - 2I^2_{at} + I^2_{pt}}{I^2_{agt}}$. Hence, (100) implies that the relative vega of the smile trade is positively proportional to this finite difference. So long as the three implied variance rates do not plot on a straight line, relative vega does not vanish. As a result, the smile trade is not a riskless arbitrage in general.

To highlight the distinction between statistical and riskless arbitrage, suppose that at some fixed time $t \in [0, T]$, the three IV's plot on a straight line.

$$I^2_{ct} - I^2_{at} = I^2_{at} - I^2_{pt}, \quad t \in [0, T].$$

(101)

Suppose further that the vol of IV is very high, e.g. $\omega_t = \infty$. Then the smile trade (74) is a statistical arbitrage and furthermore the vega of the smile trade vanishes from (78). The foreign currency delta of the smile trade can be made to also vanish by taking the appropriate position in forward contracts. However, the zero-delta zero-vega smile trade will not be a riskless arbitrage, because (101) does not imply that the three implied variance rates always plot on a straight line. The vega of the smile trade vanishes under (101), but the relative vega does not. As a result, even the delta-hedged smile trade is not a riskless arbitrage.

Now suppose instead of (101) that at some fixed time $t \in [0, T]$, the three implied variance rates do plot on a straight line, i.e.:

$$I^2_{ct} - I^2_{at} = I^2_{at} - I^2_{pt}, \quad t \in [0, T].$$

(102)

In this case, the convexity measure $c_t$ defined in (95) vanishes. Since $\omega_t > 0$, (97) holds, implying that the $Q$-drift in value of the smile trade is positive. As a result, the smile trade is a statistical arbitrage. Furthermore, the relative vega of the smile trade vanishes. Hence, when the three implied variance rates plot on a straight line as in (102), then the delta-hedged smile trade (74) is both a statistical arbitrage and a riskless arbitrage.

8 Riskless Arbitrage Under Insufficient Skew or Smile

The last section gave sufficient conditions under which the vol, skew, or smile trade enjoyed a positive $Q$-mean gain rate. To also be a riskless arbitrage in our SV model, the foreign currency delta and the relative vega of the trade must both vanish simultaneously. Given the ability to dynamically trade forward contracts, it is straightforward to delta hedge each trade without affecting the positivity of the mean gain on the trade.

When at most three strike rates can be held, it is less straightforward to zero out the relative vega, while maintaining positive $Q$-drift and staying exposed to just one of $\sigma^2_t$, $\gamma_t$, and $\omega^2_t$. In
fact, since the vol trade just contains ATM straddles, it is impossible to simultaneously zero out relative vega, maintain positive $Q_-$ drift, and stay positively exposed to just $\sigma_t^2$. In contrast, when the relative vega of the skew trade is non-zero, one can attempt to use ATM straddles to zero out the relative vega, without introducing an unwanted exposure to either of $\sigma_t^2$ and $\gamma_t$. We will give sufficient conditions under which this alteration preserves positivity of the $Q_-$ drift. Similarly, when the relative vega of the smile trade is non-zero, one can again attempt to use ATM straddles to zero out the relative vega, without introducing an unwanted exposure to either of $\sigma_t^2$ and $\omega_t^2$. We will again give sufficient conditions under which this alteration preserves positivity of the $Q_-$ drift.

Taking expectations under $Q_-$ in (50), the $Q_-$ drift of the ATM straddle is positively proportional to the difference between the instantaneous variance rate $\sigma_t^2$ and the ATM implied variance rate $I_{at}^2$:

$$G_t^v = RT_t^\sigma \left[ \frac{\sigma_t^2}{2} - \frac{I_{at}^2}{2} \right] \quad t \in [0, T].$$

(103)

Until now, we have been assuming that the random variable $\sigma_t^2$ has strictly positive variance at $t$. In contrast, the ATM IV $I_{at}$ is known at time $t$, so has zero variance at $t$. For the remainder of this paper, we assume that $\sigma_t^2$ also has zero variance and is known to all at time $t$. This new assumption is actually standard in all SV models. Let $\sigma_t > 0$ be the positive square root of $\sigma_t^2$. The next subsection examines the possible riskless arbitrage at time $t \in [0, T]$ in the simpler case when the market-maker sets $I_{at} = \sigma_t$. The following section examines the possible riskless arbitrage at time $t \in [0, T]$ in the more complicated case when $I_{at}^o \neq \sigma_t$.

### 8.1 ATM IV Equal to Instantaneous Volatility

Until now, we have allowed arbitrages in option portfolios with the exception of put call parity. In the standard SV approach, one treats $\sigma_t^2$, $\gamma_t$, and $\omega_t^2$ as observed at $t$, and then arbitrage among options is avoided by setting the level, slope, and curvature of the implied variance rate curve so as to zero out the risk-neutral drift of all option portfolio values. By instead treating $\sigma_t^2$, $\gamma_t$, and $\omega_t^2$ as random variables at $t$, as we have done until now in this paper, the option portfolio drift becomes random and hence cannot be zeroed out by the choice of IV’s.

However suppose that just $\sigma_t^2$ is observed at $t$, while $\gamma_t$ and $\omega_t^2$ remain random at $t$. In this case, the absence of arbitrage in portfolios involving ATM straddles implies in our setting that the ATM IV $I_{at}$ must always equal the instantaneous volatility $\sigma_t$. When this occurs, the ATM straddle value has zero risk-neutral drift and vega hedging using ATM straddles becomes costless. Since delta hedging using forward contracts is also costless, statistical arbitrages can be converted into riskless arbitrages costlessly. We explore the details of this conversion in this subsection. In the next subsection, we continue to assume that just $\sigma_t^2$ is observed at $t$, but we allow arbitrage in the ATM straddle quote. In that subsection, we show that a vega and delta-hedged normalized risk-reversal is an arbitrage if the slope coefficient in a regression of log IV on log spot exceeds the relative slope of implied variance. We also show that a vega and delta-hedged normalized butterfly-spread is an arbitrage if the ratio of implied volatility is greater than one. We will also develop alternative sufficient conditions which lead to alternative trading strategies being riskless arbitrages.
In this subsection only, we now assume that at some fixed time \(t \in [0, T]\), the options market-maker sets his ATM IV \(I_{at}\) equal to the known value of the instantaneous volatility \(\sigma_t\):

\[
I_{at} = \sigma_t. \tag{104}
\]

### 8.1.1 Arbitraging Insufficient Slope of Halved Implied Variance in Log Moneyness

In this subsubsection, we further assume that the investor knows that the market-maker’s slope in log moneyness of halved implied variance is insufficiently positive, given the lowest possible realization of the random covariation rate \(\gamma_t\) between the log FX rate \(S_t\) and the log IV curve \(I_t(K)\). When both (85) and (104) hold, the positive \(\mathbb{Q}^-\) drift condition (87) holds, for any choice of the ATM straddle holdings \(\eta^a_t\).

We saw in the last section that having no position in the ATM straddle will in general lead to nonzero relative vega of the skew trade that induces positive \(\mathbb{Q}^-\) drift. Recall from (89) that the relative vega of the skew trade is:

\[
\eta^a_t Rv^a_t = \eta^a_t R\Gamma^a_t I_{at}^2 \quad t \in [0, T]. \tag{105}
\]

As a result, setting:

\[
\eta^p_t = -\frac{T - t}{R\Gamma^p_t} \frac{I_{ct}^2 - I_{pt}^2}{I_{at}^2} \quad t \in [0, T], \tag{106}
\]

neutralizes relative vega at the portfolio level, while preserving positive \(\mathbb{Q}^-\) drift of the option portfolio. Combining (65) with (106), we refer to the three strike rate option portfolio:

\[
\begin{align*}
\eta^p_t &= -\frac{1}{(\ell^c_t - \ell^p_t)R\Gamma^p_t} \\
\eta^0_t &= -\frac{T - t}{(\ell^c_t - \ell^p_t)R\Gamma^0_t} \frac{I_{ct}^2 - I_{pt}^2}{I_{at}^2} \\
\eta^c_t &= \frac{1}{(\ell^c_t - \ell^p_t)R\Gamma^c_t}
\end{align*} \tag{107}
\]

as a relative vega-hedged skew trade.

To generate riskless arbitrage, one only needs to further neutralize portfolio delta using a co-terminal forward contract. Recall from (51) that one ATM straddle has the following delta:

\[
\Delta^a_t = N(I_{at}\sqrt{T - t}) - N(-I_{at}\sqrt{T - t}) \quad t \in [0, T]. \tag{108}
\]

It follows from (106) that \(\eta^a_t\) ATM straddles have the following delta:

\[
\eta^a_t \Delta^a_t = -\frac{T - t}{(\ell^c_t - \ell^p_t)R\Gamma^a_t} \frac{I_{ct}^2 - I_{pt}^2}{I_{at}^2} [N(I_{at}\sqrt{T - t}) - N(-I_{at}\sqrt{T - t})] \quad t \in [0, T]. \tag{109}
\]
Hence, when the skew trade (65) has its relative vega zeroed out, (88) implies that the foreign currency delta of this three strike rate portfolio is:

$$\Delta^v_t = \frac{1}{\ell^c_t - \ell^p_t} \left\{ \frac{N(\ell^p_{t+1})}{R^{\ell^p}_t} + \frac{N(-\ell^p_{t+1})}{R^{\ell^p}_t} \right\} - \frac{T - t}{R^{\ell^a}_t} \frac{I^2_{ct} - I^2_{pt}}{I^2_{at}} \left[ N(I_{at}\sqrt{T - t}) - N(-I_{at}\sqrt{T - t}) \right]$$

This is the number of forward contracts that an investor must short to convert the relative vega-hedged skew trade (107) into a riskless arbitrage.

We conclude that if the market-maker sets his ATM IV to the instantaneous volatility and if a market-maker has insufficient slope in log money of halved implied variance, then a delta-neutral skew trade is itself not a riskless arbitrage opportunity unless $I_{ct} = I_{pt}$ so that relative vega vanishes. When $I_{ct} \neq I_{pt}$, then the skew trade has nonzero relative vega and the trader must as a result also take a position in ATM straddles to neutralize the relative vega of the skew trade. Since these ATM straddles have nonzero delta in our SV model, the delta-hedge must account for the delta introduced by the relative vega hedge.

### 8.1.2 Arbitraging Insufficient Implied Variance Convexity in Log-moneyness

In this subsubsection, we assume that the investor has superior information on the volatility of IV, $\omega_t$, rather than on the covariation rate $\gamma_t$. As a result, we replace our assumption (85) with (96). Thus, the investor knows that the market-maker’s convexity in log-money is insufficiently positive, given the lowest possible realization of $\omega_t$. When (96) and (104) both hold, the positive Q-drift condition in (98) holds, for any choice of the ATM straddle holdings $\eta^p_t$. A nonzero position in an ATM straddle can be used to zero out the relative vega of the smile trade when this relative vega is nonzero.

Recall from (100) that the relative vega of the smile trade is $(T - t) \frac{I^2_{ct} + I^2_{pt} - I^2_{at}}{2I^2_{at}}$. If $I_{pt} = I_{at} = I_{ct}$, then the smile trade already has zero relative vega. If the three IV’s do not plot flat, then from (105), one can augment the number of ATM straddles held by:

$$- \frac{T - t}{R^{\ell^a}_t} \frac{I^2_{ct} - I^2_{at}}{2I^2_{agt}} = - \frac{T - t}{R^{\ell^a}_t} \frac{2}{2I^2_{agt}} \left[ \frac{I^2_{ct} + I^2_{pt}}{2I^2_{at}} - 1 \right] t \in [0, T].$$

This choice neutralizes relative vega at the portfolio level, while preserving positive Q-drift. Combining (74) with (111), we refer to the three strike rate option portfolio:

$$\eta^p_t = \frac{1}{I^2_{agt} R^{\ell^p}_t} \quad \eta^a_t = \frac{1}{R^{\ell^a}_t I^2_{agt}} \left\{ 2 + (T - t) \left[ \frac{I^2_{ct} + I^2_{pt}}{2I^2_{at}} - 1 \right] \right\} \quad \eta^c_t = \frac{1}{I^2_{agt} R^{\ell^c}_t}$$

as a relative vega-hedged smile trade.

To generate a riskless arbitrage opportunity, one only needs to neutralize portfolio delta using the underlying co-terminal forward contract. From (108) and (112), the foreign currency delta of the adjustment to the ATM straddle position is:

$$- \frac{1}{R^{\ell^a}_t I^2_{agt}} \left( T - t \right) \left[ \frac{I^2_{ct} + I^2_{pt}}{2I^2_{at}} - 1 \right] \left[ N(I_{at}\sqrt{T - t}) - N(-I_{at}\sqrt{T - t}) \right] t \in [0, T].$$

28
As a result, the foreign currency delta of the relative vega-hedged smile trade at time $t \in [0, T]$ is:

$$
\Delta^\gamma_t = \frac{1}{\ell^2_{agt}} \left\{ N\left(-\ell_{at}\right) - \left\{ 2 + (T - t) \left[ I^2_{at} + I^2_{pt} \right] / 2I^2_{at} - 1 \right\} \right\} \frac{N(I_{at}\sqrt{T - t}) - N(-I_{at}\sqrt{T - t})}{R\Gamma_t} - \frac{N(I^2_{pt})}{R\Gamma_t^p} \right\}.
$$

(114)

This is the number of forward contracts that an investor must short to convert the relative vega-hedged smile trade (112) into a riskless arbitrage.

We conclude that if the market-maker sets his ATM IV to the instantaneous volatility and if the market-maker has insufficient convexity of implied variances in log-moneyness, then a delta-neutral smile trade is not a riskless arbitrage unless the three IV’s equate. When they do not, the trader must adjust the holding in ATM straddles to neutralize relative vega and a position in forward contracts to neutralize the foreign currency delta.

### 8.2 ATM IV Differs from Instantaneous Vol

In this section, we deal with the more complicated situation where the options market-maker sets the value of the ATM IV $I_{at}$ different from the known value of the instantaneous volatility, $\sigma_t$:

$$
I_{at} \neq \sigma_t.
$$

(115)

We will see that the previous case where $I_{at} = \sigma_t$ is just a special case of the analysis in this section.

#### 8.2.1 Arbitraging Insufficient Relative Slope in Log-moneyness

Recall that (64) expresses the $Q_-$ mean gain rate of the option portfolio when the OTM put and call are equally OTM. To eliminate exposure to $\omega_t$, suppose we further set $\eta^R_t R\Gamma^c_t = -\eta^c_t R\Gamma^c_t$ in (64). The $Q_-$ mean gain rate of the option portfolio at time $t \in [0, T]$ is then given by:

$$
G^\gamma_t = \frac{\eta^R_t R\Gamma^a_t}{2} \left( \sigma^2_t - I^2_{at} \right) + \eta^c_t R\Gamma_t (\ell^c_t - \ell^p_t) \left( \gamma_t - \frac{I^2_{at}}{2} - \frac{I^2_{pt}}{2} \right) \equiv \mu_t, \quad t \in [0, T].
$$

(116)

We have set the drift of this portfolio to $\mu_t$. One of our objectives is to find conditions under which $\mu_t > 0$. We can treat (116) as one linear equation in the two unknowns $\eta^a_t$ and $\eta^c_t$.

To be a riskless arbitrage, it is necessary that the relative vega of the option portfolio vanishes. Setting $\eta^R_t R\Gamma^a_t = -\eta^c_t R\Gamma^c_t$ in (57), the relative vega of the option portfolio becomes:

$$
R\gamma^\gamma_t = (T - t) \left[ \eta^R_t R\Gamma^a_t I^2_{at} + \eta^c_t R\Gamma^c_t (I^2_{at} - I^2_{pt}) \right] \quad t \in [0, T].
$$

(117)

Setting this relative vega to zero gives us a second linear equation in $\eta^a_t$ and $\eta^c_t$:

$$
\eta^R_t R\Gamma^a_t I^2_{at} + \eta^c_t R\Gamma^c_t (I^2_{at} - I^2_{pt}) = 0 \quad t \in [0, T].
$$

(118)

Suppose we multiply (118) by $\frac{\sigma^2_t - I^2_{at}}{2I^2_{at}}$:

$$
\frac{\eta^R_t R\Gamma^a_t}{2} \left( \sigma^2_t - I^2_{at} \right) + \eta^c_t R\Gamma^c_t \frac{\sigma^2_t - I^2_{at}}{2I^2_{at}} (I^2_{at} - I^2_{pt}) = 0 \quad t \in [0, T].
$$

(119)
The first terms on the LHS of (116) and (119) are both \( \frac{\eta^c_t}{2} (\sigma_t^2 - I_{at}^2) \). Subtracting (119) from (116) implies:

\[
\eta^c_t R \Gamma^c_t \left[ (\ell^c_{-t} - \ell^p_{-t}) \left( \gamma_t - \frac{I_{at}^2}{\ell^c_{-t} - \ell^p_{-t}} \right) - \frac{\sigma_t^2}{I_{at}^2} \left( \frac{I_{at}^2}{2} - \frac{I_{pt}^2}{2} \right) \right] = \mu_t, \quad t \in [0, T].
\]

(120)

Simplifying the LHS:

\[
\eta^c_t R \Gamma^c_t \left[ (\ell^c_{-t} - \ell^p_{-t}) \gamma_t - \frac{\sigma_t^2}{I_{at}^2} \left( \frac{I_{at}^2}{2} - \frac{I_{pt}^2}{2} \right) \right] = \mu_t, \quad t \in [0, T].
\]

(121)

Let \( \beta_t \equiv \frac{\gamma_t}{\sigma_t^2} \) be the slope coefficient in a univariate regression of log IV on log spot. Factoring out \((\ell^c_{-t} - \ell^p_{-t})\sigma_t^2 > 0\) on the LHS of (121):

\[
\eta^c_t R \Gamma^c_t (\ell^c_{-t} - \ell^p_{-t}) \sigma_t^2 \left[ \beta_t - \frac{I_{at}^2}{2} - \frac{I_{pt}^2}{2} \right] = \mu_t, \quad t \in [0, T].
\]

(122)

Suppose that at some fixed time \( t \in [0, T] \), the relative slope measure \( \frac{I_{at}^2 - I_{pt}^2}{(\ell^c_{-t} - \ell^p_{-t})I_{at}^2} \) is less than the slope coefficient \( \beta_t \):

\[
\frac{I_{at}^2 - I_{pt}^2}{(\ell^c_{-t} - \ell^p_{-t})I_{at}^2} < \beta_t \equiv \frac{\gamma_t}{\sigma_t^2}, \quad t \in [0, T].
\]

(123)

If (123) holds, then (121) implies that \( \eta^c_t \) and \( \mu_t \) have the same sign. We henceforth assume that (123) holds and we set \( \eta^c_t > 0 \), so that \( \mu_t > 0 \). If a trader now picks a positive \( \mu_t \) in (122), then the positive number of OTM calls to hold to achieve this positive drift is given by solving (122) for \( \eta^c_t \):

\[
\eta^c_t = \frac{R \Gamma^c_t (\ell^c_{-t} - \ell^p_{-t}) \sigma_t^2 \left[ \beta_t - \frac{I_{at}^2}{2} - \frac{I_{pt}^2}{2} \right]}{\mu_t}, \quad t \in [0, T].
\]

(124)

From (118), the number of ATM straddles held at time \( t \in [0, T] \) is given by:

\[
\eta^a_t = \frac{R \Gamma^c_t I_{at}^2 - I_{pt}^2}{I_{at}^2} \eta^c_t, \quad t \in [0, T].
\]

(125)

Equation (125) implies that \( \eta^a_t \) has the opposite sign of the slope of the line connecting \( I_{pt}^2 \) to \( I_{at}^2 \). Hence, if the market-maker’s slope is positive, the ATM straddle holdings are negative and vice versa.

In conclusion, if (123) holds at some fixed time \( t \in [0, T] \), then by setting \( \eta^p_t = \frac{R \Gamma^c_t}{R \Gamma^a_t} \eta^c_t < 0, \eta^a_t \) according to (125), and \( \eta^c_t > 0 \) according to (124), a trader creates an option portfolio with positive \( \mathbb{Q}_- \) mean gain rate and zero relative vega at time \( t \in [0, T] \). A trader can then use a nonzero position \( \eta^f_t \) in the forward contract to neutralize the foreign currency delta of the three strike rate

30
option portfolio. This arbitrage portfolio can be interpreted as a normalized risk-reversal whose relative vega is zeroed out using ATM straddles. The foreign currency delta of this three strike rate portfolio is then zeroed out using forward contracts. The key point is that the sufficient condition (123) for riskless arbitrage has simple geometric and statistical interpretation. This simplifies to the slope condition (85) on just halved implied variance when \( I_{at} = \sigma_t \).

8.2.2 Arbitraging Insufficient Relative Curvature in Log-moneyness

To eliminate exposure to \( \gamma_t \) and hence \( \rho_t \), suppose we set \( \eta_t^p R\Gamma_t^p = \eta_t^c R\Gamma_t^c \) in (73). Then the \( Q_- \) mean gain rate of the option portfolio at time \( t \in [0, T] \) is given by:

\[
G_t^v = \frac{\eta_t^p R\Gamma_t^p}{2} (\sigma_t^2 - I_{at}^2) + \frac{\eta_t^c R\Gamma_t^c}{2} \left( \omega_t^2 - \frac{I_{at}^2 + I_{at}^2}{I_{at}^2} - \sigma_t^2 \right) \equiv \mu_t \quad t \in [0, T].
\]  

(126)

We treat (126) as one linear equation in the two unknowns \( \eta_t^a \) and \( \eta_t^c \).

Setting \( \eta_t^p R\Gamma_t^p = \eta_t^c R\Gamma_t^c \) in (57), the relative vega of the option portfolio becomes:

\[
Rv_t^v = (T - t)[\eta_t^a R\Gamma_t^a I_{at}^2 + \eta_t^c R\Gamma_t^c (I_{pt}^2 + I_{ct}^2)] \quad t \in [0, T].
\]  

(127)

Setting this relative vega to zero gives us a second linear equation in \( \eta_t^a \) and \( \eta_t^c \):

\[
\eta_t^a R\Gamma_t^a I_{at}^2 + \eta_t^c R\Gamma_t^c (I_{pt}^2 + I_{ct}^2) = 0 \quad t \in [0, T].
\]  

(128)

Suppose we multiply (128) by \( \frac{\sigma_t^2 - I_{at}^2}{2I_{at}^2} \):

\[
\frac{\eta_t^a R\Gamma_t^a}{2} (\sigma_t^2 - I_{at}^2) + \frac{\eta_t^c R\Gamma_t^c}{2} (\sigma_t^2 - I_{at}^2) \frac{I_{pt}^2 + I_{ct}^2}{I_{at}^2} = 0 \quad t \in [0, T].
\]  

(129)

Subtracting (129) from (126) implies:

\[
\eta_t^c R\Gamma_t^c \left[ \frac{I_{pt}^2 + I_{ct}^2}{2I_{at}^2} \left( \omega_t^2 - \frac{I_{at}^2 + I_{at}^2}{2I_{at}^2} - \sigma_t^2 \right) - (\sigma_t^2 - I_{at}^2) \frac{I_{pt}^2 + I_{ct}^2}{2I_{at}^2} \right] = \mu_t \quad t \in [0, T].
\]  

(130)

Simplifying the LHS:

\[
\eta_t^c R\Gamma_t^c \frac{\omega_t^2 - \sigma_t^2}{I_{at}^2} = \mu_t, \quad t \in [0, T].
\]  

(131)

Let \( R_t \equiv \frac{\omega_t}{\sigma_t} \) be the ratio of vol-vol to spot vol. Factoring out \( \frac{I_{pt}^2 + I_{ct}^2}{2I_{at}^2} > 0 \) on the LHS of (131):

\[
\eta_t^c R\Gamma_t^c \frac{I_{pt}^2 + I_{ct}^2}{2I_{at}^2} \left( R_t^2 - \frac{I_{pt}^2 + I_{ct}^2}{2I_{at}^2} - I_{at}^2 \right) = \mu_t, \quad t \in [0, T].
\]  

(132)
The fraction subtracted from $R_t^2$ is a measure of relative convexity of implied variances. Suppose that at some fixed time $t \in [0, T]$, this relative convexity measure is less than the variance ratio $R_t^2$,

\[
\frac{\frac{I_{at}^2 + I_{ct}^2}{2} - I_{ct}^2}{\frac{I_{at}^2}{2}} < R_t^2 \equiv \frac{\omega_t^2}{\sigma_t^2}, \quad t \in [0, T].
\]

(133)

In this case, (132) implies that $\eta_t^c$ and $\mu_t$ have the same sign. We henceforth assume that (133) holds and that we set $\eta_t^c > 0$ so that $\mu_t > 0$. If a trader now picks a positive $\mu_t$ in (132), then the positive number of OTM calls to hold to achieve this positive drift is given by solving (132) for $\eta_t^c$:

\[
\eta_t^c = \frac{\mu_t}{R \Gamma_t^c \Gamma_t^a \sigma_t^2} \left[ R_t^2 - \frac{I_{at}^2 + I_{ct}^2 - I_{ct}^2}{\frac{I_{at}^2}{2}} \right], \quad t \in [0, T].
\]

(134)

From (128):

\[
\eta_t^a = \frac{R \Gamma_t^c}{R \Gamma_t^a} \frac{I_{pt}^2 + I_{ct}^2}{2I_{at}^2} \eta_t^c, \quad t \in [0, T],
\]

(135)

where $\eta_t^c$ is given in (134). Since the sign of $\eta_t^c$ is positive, the sign of $\eta_t^a$ is negative.

In conclusion, if (133) holds at some fixed time $t \in [0, T]$, then by setting $\eta_t^p = \frac{R \Gamma_t^c}{R \Gamma_t^a} \eta_t^c > 0$, $\eta_t^a < 0$ according to (135), and $\eta_t^c > 0$ according to (134), one can have an option portfolio with positive $\mathbb{Q}_-$ mean gain rate and zero relative vega. One can then use a nonzero position $\eta_t^f$ in the forward contract to neutralize the delta of the option portfolio. This arbitrage portfolio can be interpreted as a normalized strangle whose relative vega is zeroed out using ATM straddles. The resulting foreign currency delta of the three strike rate option portfolio is then zeroed out using forward contracts. The key point is that the sufficient condition (133) for arbitrage when $I_{at} \neq \sigma_t$ has simple geometric and statistical interpretation. This simplifies to the convexity condition (96) on implied variance when $I_{at} = \sigma_t$.

9 Summary and Future Research

We considered an options market where a market-maker continuously quotes the continuum of BMS IV’s by strike rate for some fixed maturity date. We further supposed that the IV curve had the same type of $\mathbb{Q}_-$ dynamics as the spot FX rate, namely driftless geometric Brownian motion generalized to have arbitrary unspecified volatility.

In this setting, we showed that in a partially unspecified SV setting, an ATM option can be dynamically traded to profit on average from the difference between the instantaneous variance rate of log FX and the ATM option’s Black Merton Scholes (BMS) implied variance rate. We also showed that a pair of options can be dynamically traded to profit on average from the difference between the log FX - log IV covariation and half the slope of a line connecting two points on a BMS implied variance curve. Finally, we showed that an option triple can be dynamically traded to profit on average from the difference between the instantaneous variance rate of log IV and
a standard convexity measure of the BMS implied variance curve. Our results yielded precise financial interpretations of particular measures of the level, slope, and curvature of an IV curve. These interpretations help explain standard quoting conventions found in the OTC market for options written on precious metal and foreign exchange.

We then gave sufficient conditions under which an investor with superior information can create a statistical arbitrage against a market-maker who has too low an ATM IV, insufficient slope between the two OTM halved implied variance quotes, or insufficient convexity. While no jumps and non-trivial relative shifts were assumed for both FX and IV, the portfolio construction does not require knowledge of the initial level or dynamics of the three stochastic processes $\sigma$, $\omega$, and $\rho$ generating volatilities and correlations.

We then investigated when riskless arbitrage was possible. For the skew trade, a special case arises when the OTM IV’s match. In this case the statistical arbitrage reduces to a long position in a risk-reversal, which will have zero relative vega. As a result, the delta-hedged long position in the risk-reversal becomes a riskless arbitrage.

When the OTM IV’s do not match, we assumed that the instantaneous volatility $\sigma_t$ is not random and observed at $t$, while we continued to assume that $\omega_t$ and $\rho_t$ are random at $t$. We first investigated the simpler case when the absence of arbitrage in ATM straddle quotes causes the market-maker to set his ATM IV to $\sigma_t$. In this case, the same conditions that lead to statistical arbitrage also lead to riskless arbitrage. A trader merely has to zero out both the relative vega and the foreign currency delta of the skew and smile trade using ATM straddles and forward contracts respectively.

We then investigated the more complicated case when the market-maker does not set his ATM IV to $\sigma_t$. We showed that when the random covariation rate $\gamma_t \equiv \rho_t \sigma_t \omega_t$ is certain to exceed the slope measure $\frac{I_2^c}{I_2^p} \frac{\sigma_t^2}{\rho_t \omega_t}$, then a skew trade is a riskless arbitrage provided that it is relative vega-hedged using ATM straddles and delta-hedged using forward contracts.

We also showed that when a linear combination of two convexity measures is insufficiently positive, then a trade involving long OTM calls and puts is again a riskless arbitrage provided that it is relative vega-hedged using ATM straddles and delta-hedged using forward contracts. One can still refer to this trade as a smile trade, but it is worth mentioning that the positive $Q_-$ drift condition depends on both the volvol $\omega_t$ and the volatility of $S$, viz $\sigma_t$. The latter enters through our requirement that the trade both earn positive $Q_-$ drift and have zero relative vega.

Future research should explore the complications that arise when the IV update process is allowed to drift under $Q_-$. When the common IV drift is an additional unknown stochastic process, it may be necessary to replace the ATM straddle with two near-the-money options for which different positions are allowed. Alternatively, one can add an ATM straddle at a second maturity. If the IV process is chosen to mean revert, one can attempt to calibrate the speed of mean reversion process and the long run mean process by introducing two more ATM straddle maturities to the current setup.

Future research should investigate the use of alternative measures for the attractiveness of a trade. Positive $Q_-$ drift is too weak a measure since there is no control over the risk. Riskless arbitrage is too strong a measure since the resulting Sharpe Ratio is infinite. One can try to design
a Sharpe Ratio criterion under $\mathbb{P}$ or one can alternatively just demand that the real-valued random drift be “acceptable”.

Future research should also investigate the robustness of the statistical and riskless arbitrage portfolios to jumps. In the riskless arbitrage case, one is interested in the size of the neighborhood of $(W,Z)$ for which the dynamically traded option portfolio has weakly greater value$^2$. Future research should also investigate the informational requirements of a one factor model, rather than a model with constant relative shifts across all IV’s. In the interests of brevity, these extensions are best left for future research.

---

$^2$This is related to Bregman dispersion.
References


