# Section 6 Implied Volatility and the Smile

Financial Engineering · Fall 2026 · Zong-Wei Yeh · Department of Finance, CCU

## Overview

Section 4 built the Black–Scholes model. This section compares it with market prices. The tool is implied volatility: the volatility that makes the Black–Scholes formula reproduce a quoted price (§6.1). Section 6.2 explains why markets and researchers work with this number rather than with prices. Section 6.3 reads implied volatilities in real TXO data (options on the TAIEX index). They are not constant across strikes and maturities. The pattern shows what a better model needs. Section 6.4 proves the Breeden–Litzenberger formula, which reads the risk-neutral density off option prices. Section 6.5 puts all strikes and maturities together into the implied volatility surface. It states the surface's no-arbitrage conditions and builds one from TXO quotes. Section 6.6 reads five option positions as trades on the surface. It also tests the straddle as a proxy for the variance risk premium on TXO data. Section 6.7 explains the VIX: the old index, the idea behind the new formula, and what the index means today. Section 6.8 states Dupire's local volatility formula and explains it. Section 6.9 shows four ways the surface is used. Dupire's formula is stated with its intuition and a reference. The other results are derived in short steps. Figures 1, 2 and 5 use the instructor's TXO research panel. Figure 3 uses index data of the Cboe (Chicago Board Options Exchange). Figure 9 uses TAIFEX settlement prices and TWSE index closes. Figure 10 uses the panel and TWSE index closes. Every data figure is made by a committed script in `figures/`.

### 6.1 Implied Volatility

This subsection defines implied volatility and shows that it is well defined. In the Black–Scholes formula every input but one is observed. We know $S$, $K$, $r$ and $\tau$, but not $\sigma$. So we turn the question around. We ask which $\sigma$ makes the formula match the market price.

**Definition 6.1** (Implied volatility)**.**
*Take a quoted call price $C^{\text{mkt}}$ with strike $K$ and time to maturity $\tau$, and the observed $S$ and $r$. The* implied volatility $\sigma_{\text{imp}}(K,\tau)$ *is the $\sigma$ that solves*

$$C^{\text{BS}}\big(S,K,r,\tau;\sigma\big) \;=\; C^{\text{mkt}}.$$

*If the underlying pays a dividend yield $\delta$ (Section 4 writes it $q$), $C^{\text{BS}}$ is the Black–Scholes price with yield $\delta$. That is the formula with $S$ replaced by $Se^{-\delta\tau}$. A put's implied volatility is defined in the same way, with the Black–Scholes put price.*

The definition says "the" $\sigma$. So we must check that exactly one solution exists. The idea is simple: the price rises with $\sigma$, and it runs between the two no-arbitrage bounds.

**Proposition 6.2** (Vega).
*Let $\delta = 0$. The Black–Scholes call price is strictly increasing in $\sigma$:*

$$\mathcal{V} := \frac{\partial C^{\mathrm{BS}}}{\partial \sigma} = S\,\varphi(d_1)\sqrt{\tau} > 0.$$

*As $\sigma$ runs over $(0, \infty)$, $C^{\mathrm{BS}}$ takes every value in the interval $\big((S - Ke^{-r\tau})^+,\ S\big)$ exactly once.*

More volatility always makes the call worth more. The interval is the range that no-arbitrage allows for any call, in any model. With no dividends, the call pays at most $S_T$, so $C \le S$. It pays at least $S_T - K$ and at least 0, so $C \ge \max(S - Ke^{-r\tau}, 0)$. So every quote strictly inside these bounds has exactly one implied volatility. With a dividend yield $\delta$, replace $S$ by $Se^{-\delta\tau}$ throughout.

*Proof.*

□

## 6.2 Why Implied Volatility Matters

This subsection gives three reasons why traders and researchers work with implied volatility rather than with prices.

*First, it is a quoting convention.* A price in index points depends on the strike, the maturity and the level of the index. Implied volatility removes all three. For example, take the 2-day and the 261-day TXO options at strike 23000 on 2024-07-01. In the data the 2-day call and put cost 143 and 124 points. The 261-day call and put cost 1310 and 1290 points. So the prices differ by a factor of nine to ten. The implied volatilities are 19.6% and 16.9%. These two numbers can be compared directly. So traders quote and negotiate options in volatility points. They compare contracts across strikes, maturities and days in this one unit. (At 261 days the data file gives 15.2% for the call and 18.4% for the put. Those numbers use an interest rate but no dividend yield. With the forward that put–call parity gives, both prices give about 16.9%. That forward is close to the index level: $F_T/S \approx 0.998$, with $S$ the index close (§6.5). Exercise 6.3 shows why.)

*Second, it is a model diagnostic.* Suppose the Black–Scholes model were true. Then every option would return the same $\sigma$, whatever its strike and maturity. So any shape in $\sigma_{\mathrm{imp}}(K, \tau)$ rejects the model. The place where the shape appears tells us how the model fails. Section 6.3 reads the shapes.

*Third, it measures how expensive an option is.* An option is expensive when the volatility it prices is high compared with the volatility the underlying will deliver. So the research literature compares implied volatility with realized volatility.

**Definition 6.3** (Realized volatility)**.**
*Let $r_1, \ldots, r_n$ be the daily log returns of the underlying over $n$ trading days. The* realized volatility *over these days is*

$$\hat{\sigma}_{\mathrm{RV}} = \sqrt{\frac{252}{n} \sum_{i=1}^{n} r_i^2} \,.$$

The realized volatility is the annualized volatility that the underlying actually delivered. The factor 252 is the number of trading days in a year. The formula sets the mean return to zero. At a daily horizon the mean is tiny compared with the spread of returns, and it is estimated very poorly. So dropping it changes little and removes a noisy estimate.

Implied volatility is about the future. Realized volatility is about the past. There are two questions here. First, is implied volatility a good forecast? Christensen and Prabhala (1998) find that it forecasts future realized volatility better than past realized volatility does. Second, is the gap between the two a price? Goyal and Saretto (2009) sort stocks by the gap between historical realized volatility and at-the-money implied volatility. Options are cheap where the gap is large and positive. They are expensive where it is large and negative. Buying the cheap options and selling the expensive ones earns significant average returns. So the gap measures expensiveness. In their sample, trading did not remove it.

For the index the gap has a sign on average. Implied volatility is on average above the volatility later realized. This gap is the *variance risk premium*. Bollerslev, Tauchen and Zhou (2009) relate it to expected stock returns. Section 6.7 defines it exactly. Section 6.6 shows how straddle returns are used as a proxy for it. Exercise 6.1 computes a realized volatility and compares it with an implied one.

## 6.3 The Smile in the Data, and What It Demands

This subsection reads implied volatilities in real data and asks what they demand of a model. A few words describe the shapes. A *smile* is the curve of implied volatility against strike at one maturity. A *smirk* is a smile that rises toward low strikes: low strikes carry higher implied volatility than high strikes. The two ends of the curve are called the *wings*. The slope of the smile is called the *skew*.

The data are TXO options from the instructor's research panel. We compare an ordinary Monday, 2024-07-01, with 2024-08-05, a crash day. The TAIEX closed at 23,058.57 on 2024-07-01 and at 19,830.88 on 2024-08-05. So the index was 14% lower on the crash day. The TXO 30-day model-free implied volatility (`MF_Vol_30`; §6.7 explains it) jumped from 29% on 2024-08-02 to 46% on 2024-08-05. We compute every implied volatility in Figures 1, 2 and 5 ourselves from the option prices. We use Black's formula at the forward that put–call parity gives for each maturity (step 2 of §6.5).

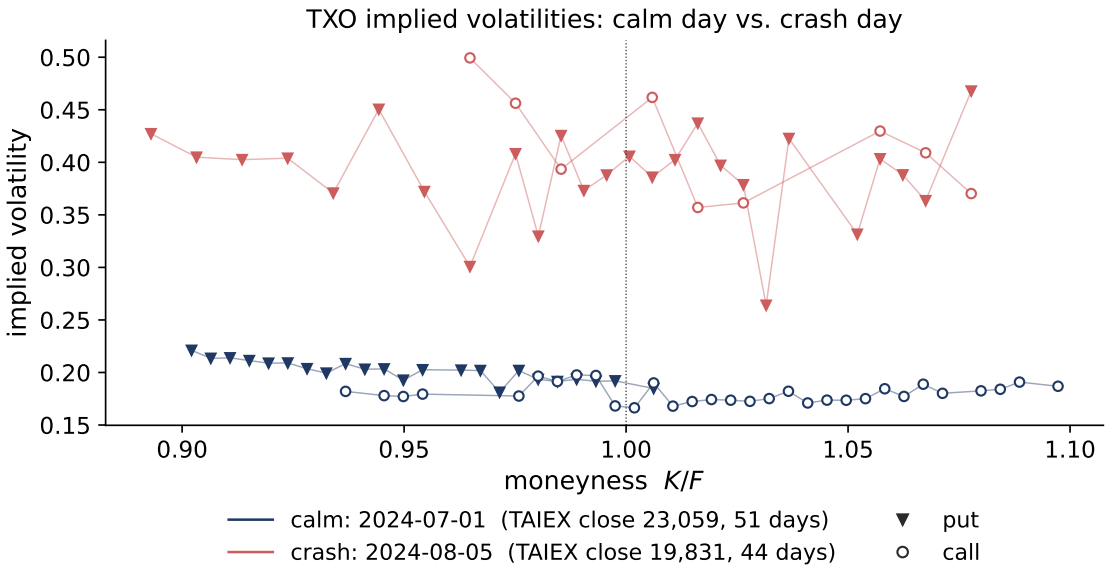



Figure 1: TXO implied volatilities by strike. Triangles are puts and circles are calls. Each type is shown at every strike with a price of at least 1 point and open interest of at least 20. Calm day: the 51-day contract. Crash day: the 44-day contract. Stale-quote outliers are filtered. Moneyness is $K/F$, with $F$ the parity forward of the contract. Calm day: the out-of-the-money puts rise toward low strikes, a smirk. Puts at 90% moneyness carry 22.1%. The at-the-money value by the rule of Figure 2 is 17.8%. At the same strike, the put's implied volatility is often about 2.5 points above the call's. Put–call parity says the two should be equal (Exercise 6.3). So these quotes do not satisfy parity exactly. A smile therefore uses one quote per strike, the out-of-the-money one (step 1 of §6.5). Crash day: both types are far higher. Most puts are at 33–43%, and most calls at 36–47%. Neighboring quotes disagree much more than on the calm day. So neither type shows a clear shape. On the crash day the slope between moneyness 0.90 and the money is not well measured. Put–call parity places the forwards of that day between 0.970 and 0.989 times the index close. The slope changes a lot within that range (Exercise 6.7).

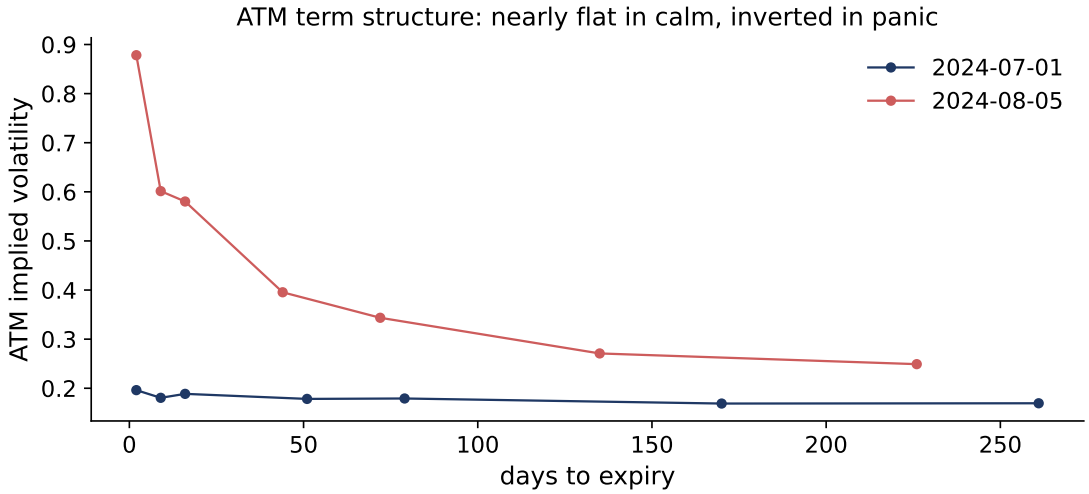



Figure 2: At-the-money implied volatility across maturities. Each value is the mean of two implied volatilities. They are the call's and the put's at the strike nearest the parity forward of that maturity. Calm day: nearly flat, between 16.9% and 19.6%. Crash day: the 2-day option trades at 87.8% and the 226-day option at 24.9%. The term structure is *inverted*. The market prices the stress as severe but short-lived.

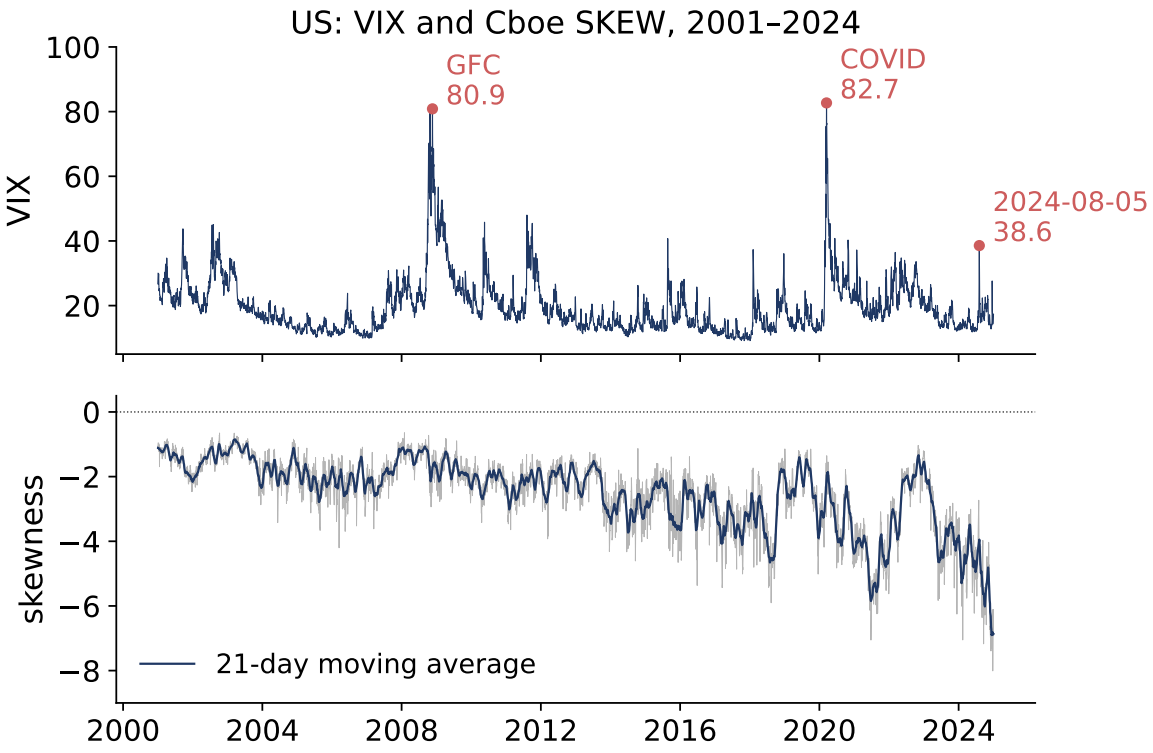



Figure 3: US data, daily closes, 2001–2024. Top: the Cboe VIX, the 30-day implied volatility of the S&P 500 in percent (§6.7 defines it). Bottom: the 30-day risk-neutral skewness of the S&P 500 log return. It comes from the Cboe SKEW index: $\text{SKEW} = 100 - 10\,S$, with $S$ the skewness (Cboe, 2010). Both series are computed from S&P 500 option prices across strikes. A negative skewness means a fat left tail. A fat left tail makes out-of-the-money puts expensive. The VIX is between 13.8 and 23.0 on half of the days. It spikes in crises: 80.9 on 2008-11-20, 82.7 on 2020-03-16, 38.6 on 2024-08-05. The skewness is negative on every day. Its median is $-2.23$. It is less negative when the VIX is high. Its median is $-1.80$ on the fifth of days with the highest VIX and $-2.69$ on the fifth with the lowest. It also drifts down over the period. Its median is $-1.40$ in 2001 and $-4.81$ in 2024. Its lowest value is $-8.0$, on 2024-12-24. Section 6.4 shows why option prices contain this information.

Four stylized facts stand out. Each one rejects a part of the Black–Scholes model.

1. **The smirk is almost permanent.**
   In the S&P 500 option market the 30-day risk-neutral skewness is negative on every trading day of 2001–2024 (Figure 3). So out-of-the-money puts usually carry higher implied volatility than out-of-the-money calls. This rejects the lognormal distribution: the market prices extra crash risk into the left tail. On the calm day of Figure 1 the out-of-the-money TXO puts show the same smirk. The crash-day slope is not well measured (Figure 1, Exercise 6.7). The robust facts of that day are the much higher level and the inverted term structure. Bates (1991) finds the same pattern in S&P 500 futures options. There, out-of-the-money puts became unusually expensive in the year before the 1987 crash.

2. **The level spikes in stress and then falls back.**
   The VIX is between 13.8 and 23.0 on half of the days. In the two largest crises it rises above 80 within weeks. Then it falls back (Figure 3, top). So constant $\sigma$ is rejected. Volatility behaves

like a mean-reverting random process, such as the Ornstein–Uhlenbeck process of Section 3.

3. **The term structure inverts in stress.**
   It is nearly flat in calm times. On 2024-07-01 it stays between 16.9% and 19.6%. It goes from 19.6% at 2 days to about 16.9% at 170 and 261 days. It is steeply inverted in panic (Figure 2). This fits mean-reverting variance. After a shock, the expected average variance over a long horizon is pulled back toward its long-run level.

4. **Short-dated smiles are the most extreme.**
   The smile is steepest at the shortest maturity. On the calm day, take the slope of implied volatility against log-moneyness at the money. It is about $-0.6$ at 2 days and $-0.23$ to $-0.33$ at 9 to 51 days. From 79 days on it is $-0.13$ or smaller in size. Section 6.5 defines log-moneyness; Figure 5 shows the slopes. This is hard for a model with continuous paths. Over two days a diffusion moves by about $\sigma\sqrt{\Delta t}$, so a large two-day move gets very little probability. A jump gives such a move directly (Section 7).

The facts now give a model specification. The argument uses the two ends of the term structure. At the short end, a diffusion moves by about $\sigma\sqrt{\Delta t}$ over a short time $\Delta t$. So large moves over a short horizon are almost impossible, and a diffusion cannot price a fat short-dated tail. A jump can. Suppose jumps arrive at rate $\lambda$ per year. Then a jump comes within a short time $\Delta t$ with probability about $\lambda\,\Delta t$. It moves the price by an amount that does not shrink with the horizon. At the long end, jumps average out. Over many periods the independent jump moves add up. So the skewness they create shrinks as the horizon grows. At 170 to 261 days the TXO skew is small, but it is not zero (Figure 5). One way to keep some skew at that horizon is a variance that is itself random. Its shocks must be negatively correlated with returns ($\rho < 0$). Facts 2 and 3 add that the variance mean-reverts. The crash day illustrates the negative correlation. On that day a fall in the index came with a jump in implied volatility. So a model of these data can use diffusion, jumps, and random variance with negative correlation. This is Bates' model. Section 7 builds it from Merton's jumps and Heston's variance.

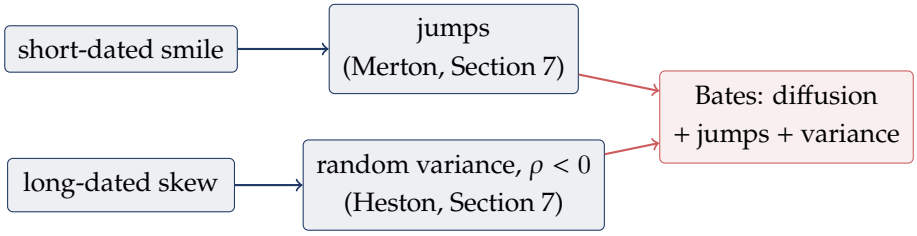



Figure 4: What each end of the term structure needs. The short end needs jumps. The small long-end skew can come from random variance with negative correlation. The two together are Bates' model.

## 6.4 The Breeden–Litzenberger Formula

This subsection extracts the whole risk-neutral distribution from a set of option prices, with no model. It is the tool behind both panels of Figure 3, including the VIX.

**Theorem 6.4** (Breeden–Litzenberger)**.**
*Let $C(K)$ be the time-0 prices of calls of all strikes $K$ with a fixed maturity $T$. Let $q$ be the risk-neutral density of $S_T$, continuous at $K$. Then*

$$\frac{\partial C}{\partial K} = -e^{-rT}\,\boldsymbol{Q}(S_T > K), \qquad \frac{\partial^2 C}{\partial K^2} = e^{-rT}\,q(K).$$

The slope of the call price in the strike gives the probability of exercise. Its curvature gives the density itself. So call prices at many strikes determine the whole risk-neutral distribution. Section 1 said that prices are expectations under $\boldsymbol{Q}$. This formula goes the other way: from prices to the distribution. The result is due to Breeden and Litzenberger (1978).

*Proof.*

□

The second derivative has a discrete version that trades. A *butterfly spread* is long one call at $K - h$, short two at $K$, and long one at $K + h$. Its price is a second difference:

$$C(K-h) - 2C(K) + C(K+h) \;\approx\; e^{-rT} q(K)\, h^2.$$

So a butterfly is a bet on the density near $K$. Its payoff is never negative, so its price must not be negative either. In other words, $C(K)$ must be convex in $K$. This is a model-free no-arbitrage condition. It is the smile's version of Section 1's rule that state prices are positive. Here $e^{-rT} q(K)\, dK$ is the state price of the states near $K$.

Integrals of option prices across all strikes give other model-free numbers about $q$. The VIX and the risk-neutral skewness of Figure 3 are such integrals of S&P 500 option prices. Section 6.7 builds the best known one, the VIX.

## 6.5 The Implied Volatility Surface

This subsection puts all strikes and maturities of one day together. It names the coordinates, states which shapes are free of arbitrage, and builds a surface from TXO quotes.

Throughout, $r$ and the dividend yield $\delta$ of Definition 6.1 are constants. (Section 4 writes the dividend yield as $q$. Here $q$ is the density, so the yield is $\delta$.) The forward price for delivery at $T$ is $F_T = Se^{(r-\delta)T}$.

**Definition 6.5** (Implied volatility surface)**.**
*The* log-moneyness *of strike $K$ at maturity $T$ is $k = \ln(K/F_T)$. The* implied volatility surface *is the function $(k, T) \mapsto \sigma_{\text{imp}}(k, T)$. The* total implied variance *is*

$$w(k, T) \;=\; \sigma_{\text{imp}}^2(k, T)\, T.$$

The coordinate $k$ measures how far the strike is from the forward, in percent. For example, $k = -0.10$ is a strike about 10% below the forward. The total implied variance $w$ is the variance of $\ln S_T$ that the option prices. The volatility $\sigma_{\text{imp}}$ is a rate per year. The variance $w$ is the total amount up to $T$. Across maturities, $w$ is the quantity that must behave well.

There are several ways to measure moneyness. Figure 1 uses $K/F_T$. The surface uses its logarithm, $\ln(K/F_T)$. Both divide by the forward, not the spot. The reason is that the forward is the center of the risk-neutral distribution at $T$. Some desks use $K/S$. Some desks divide by $\sigma\sqrt{T}$, so that one unit means one standard deviation at every maturity. Over-the-counter currency markets quote by the option's delta. All of them serve the same purpose: they make strikes comparable across maturities.

**Lemma 6.6** (Black–Scholes in surface coordinates)**.**
*Write the normalized call price as $c = C/(e^{-rT} F_T)$. Under Black–Scholes with dividend yield $\delta$,*

$$c \;=\; \Phi(d_1) - e^{k}\,\Phi(d_2), \qquad d_{1,2} = -\frac{k}{\sqrt{w}} \pm \frac{\sqrt{w}}{2}.$$

*So $c$ depends only on $k$ and $w$, and it is strictly increasing in $w$.*

The normalized price $c$ is the call price in units of the discounted forward. So moneyness and total variance are the only two numbers that matter for a call. A higher $w$ always means a higher price. We write $g(k, w) = \Phi(d_1) - e^{k}\Phi(d_2)$ for this function of two numbers.

*Proof.*

□

**Proposition 6.7** (Static no-arbitrage on the surface)**.**
*Suppose call prices are risk-neutral expectations, $C(K,T) = e^{-rT}\mathbb{E}^{Q}[(S_T - K)^+]$, under a measure $Q$ for which $e^{-(r-\delta)t}S_t$ is a martingale. Then:*

*(a)* Butterfly condition. *At each $T$, $K \mapsto C(K,T)$ is convex.*

*(b)* Calendar condition. *At each $k$, $T \mapsto w(k,T)$ is non-decreasing.*

Part (a) says that every butterfly has a non-negative price. Part (b) says that, at fixed moneyness, a longer option never prices less total variance. Part (b) is about $w$, not about $\sigma_{\text{imp}}$. So an inverted implied volatility curve is not by itself an arbitrage. Only a falling $w$ is. On the crash day, the at-the-money $w$ of Figure 2 rises at every maturity. It goes from 0.0042 at 2 days to 0.0384 at 226 days. So at the money, the inverted curve of Figure 2 does not break the calendar condition. Exercise 6.9 looks at one fixed strike, 19200. There, the 44-day and 72-day quotes give a $w$ that falls.

*Proof.*

□

**Building a surface from quotes.** A market gives quotes at a few strikes and maturities. A surface must be defined everywhere. There are five steps.

1. *Clean.* Keep out-of-the-money options only: puts below the forward and calls above it. Drop very cheap, thinly held and stale quotes. These are the filters of Figure 5. Figure 1 uses the same filters. It shows puts and calls at every strike that passes them.

2. *Find the forward.* Put–call parity, $C - P = e^{-rT}(F_T - K)$, gives $F_T = K + e^{rT}(C - P)$ at each strike. We take the median of these values over the strikes within 2% of a reference level, with $r = 1.5\%$. The reference level is the one in the data file. It comes from 2-day put–call parity, not from the index close. Now let $S$ be the official TAIEX close. On 2024-07-01 the close is 23,058.57, and $F_T/S$ lies between 0.997 and 1.000. On 2024-08-05 the close is 19,830.88. The forwards lie between 19,240 (2 days) and 19,611 (135 days). So $F_T/S$ lies between 0.970 and 0.989. On the crash day the option-implied forwards sit 1–3% below the index close. The gap does not grow with maturity. So it is not carry. The data do not show the cause of the gap.

3. *Smooth each maturity.* A widely used choice is the raw SVI ("stochastic volatility inspired")

parameterization (Gatheral and Jacquier, 2014):

$$w(k) = a + b\left\{\rho\,(k-m) + \sqrt{(k-m)^2 + s^2}\right\}.$$

Each parameter controls one shape. $a$ sets the overall level of variance. $b$ sets how steep the wings are. $\rho$ tilts the curve and creates the skew. $m$ shifts it left or right. $s$ rounds the bottom near the money. Here $\rho$ is only a shape parameter. It is not the correlation of §6.3. We use SVI as a smoothing tool and do not study it.

4. *Join the maturities.* At each $k$, interpolate linearly in $w$ between the two nearest maturities. Linear interpolation keeps the calendar condition (Exercise 6.8).

5. *Check.* The fitted prices must be convex in $K$, and the fitted $w$ curves must not cross.

Raw prices do not pass the check. Take the liquid out-of-the-money TXO options in the file (price at least 1 point, open interest at least 20). Look at each set of three neighboring strikes. On the calm day, 19 of 207 such triples break convexity by more than 5 index points. On the crash day, 36 of 89 do. A surface built from raw prices would therefore contain butterfly arbitrage. This is why step 3 smooths.

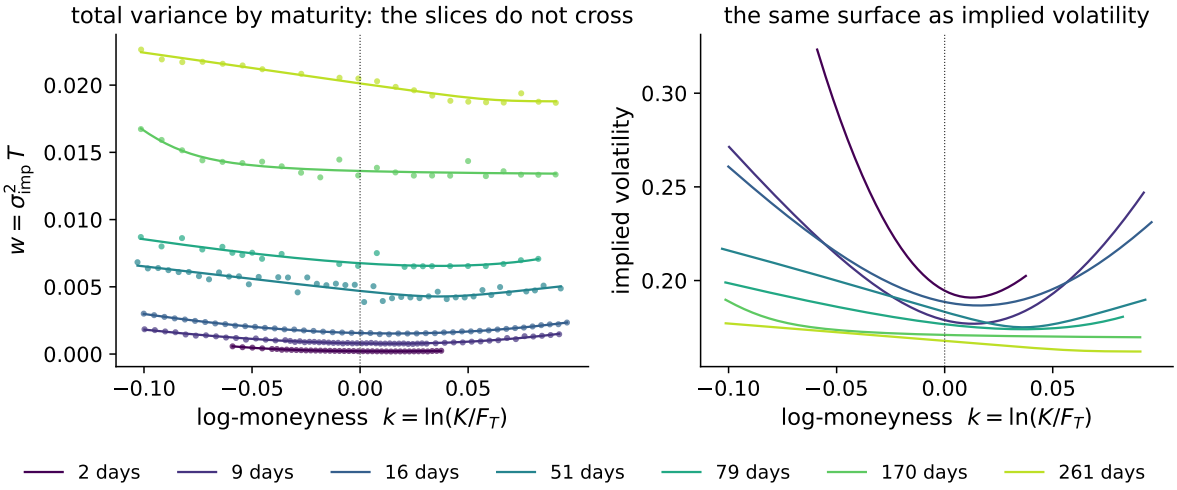



Figure 5: The TXO implied volatility surface on 2024-07-01, all seven maturities from 2 to 261 days. Left: total implied variance against log-moneyness $k = \ln(K/F_T)$, with the parity forward of each maturity. Dots are quotes and curves are SVI fits, each drawn over its own range of quotes. The fit errors are 0.1 to 0.6 volatility points. The curves do not cross for $k$ between $-0.10$ and $0.09$. So the calendar condition holds there. (The 2-day and 79-day fits are extrapolated outside their quotes.) Right: the same surface as implied volatility. The at-the-money slope $\partial\sigma_{\text{imp}}/\partial k$ is about $-0.6$ at 2 days and $-0.23$ to $-0.33$ at 9 to 51 days. It is $-0.13$ at 79 days and $-0.03$ to $-0.09$ at 170 to 261 days. So the skew flattens with maturity, but it does not vanish. The crash day is not drawn. It has no out-of-the-money puts at 2 days and only two at 9 days. Its fit errors are 1.5 to 4.4 points, and its 72-day slice has only four quotes.

## 6.6 Trading Volatility with Options

This subsection reads five option positions as trades on the surface of §6.5. Each one takes a position in one feature: the level, the tails, the skew or the term structure. It then shows why

straddle returns are used as a proxy for the variance risk premium of §6.2.

The payoffs of these positions are taught in the Options course. Here we take the theory angle. We read each position in the coordinates $(k, w)$ of Definition 6.5, and we measure its exposure with vega. Throughout, $r$ and the dividend yield $\delta$ are constants, and $F_T = Se^{(r-\delta)T}$.

**Definition 6.8** (Five volatility positions).
*All options are European, on the same underlying, with maturity $T$ unless stated otherwise. Let $K_1 < K_2$.*

*(a)* *A* straddle *at $K$ is long one call and one put, both with strike $K$.*

*(b)* *A* strangle *is long one put with strike $K_1$ and one call with strike $K_2$.*

*(c)* *A* butterfly *at $K$ with width $h$ is long one call at $K-h$, short two calls at $K$ and long one call at $K+h$ (§6.4).*

*(d)* *A* risk reversal *is long one call with strike $K_2$ and short one put with strike $K_1$.*

*(e)* *A* calendar spread *at log-moneyness $k$ is long one call with maturity $T_2$ and short one call with maturity $T_1 < T_2$. The strikes are $K_i = F_{T_i}e^k$, so both calls have the same $k$.*

In the strangle and the risk reversal the two strikes usually lie on either side of the forward, $K_1 < F_T < K_2$. If $r = \delta$, then $F_{T_1} = F_{T_2} = S$. So a calendar spread with one strike for both maturities then has a common $k$. Figure 6 draws the five positions.

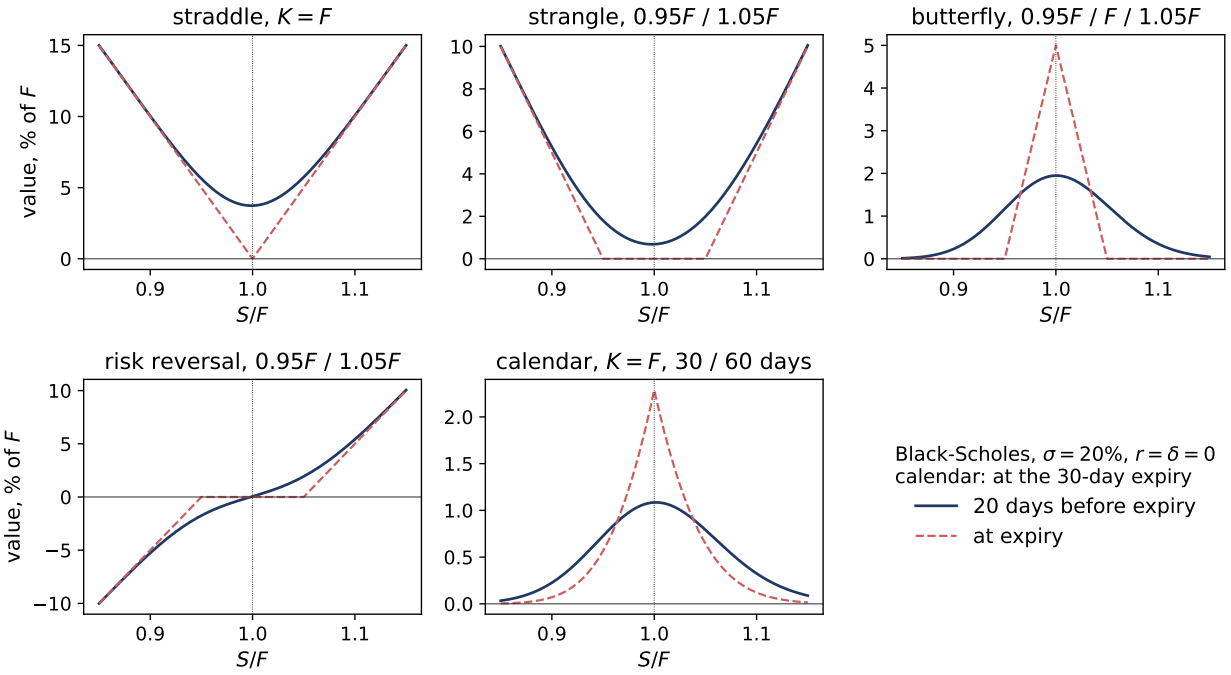



Figure 6: The five positions of Definition 6.8, valued by Black–Scholes with $\sigma = 20\%$ and $r = \delta = 0$, against $S/F$. The first maturity is 30 days. Solid: value 20 days before the first expiry. Dashed: value at the first expiry. The calendar spread ($K = F$, 30 and 60 days) is drawn at the 30-day expiry, when it still holds the 60-day call. The straddle and the strangle gain from a large move in either direction. The butterfly and the calendar spread gain when the index stays near $F$. The risk reversal gains when the index rises and loses when it falls.

**Vega on the surface.** A position is exposed to the points of the surface where its options sit. Vega measures this exposure. It is the change in an option price per unit change in $\sigma$. Take the strike $K = F_T e^k$ and put $w = \sigma^2 T$. Then the call and the put with this strike have the same vega:

$$\mathcal{V}(k) \;=\; e^{-rT} F_T \sqrt{T}\, \varphi(d_1), \qquad d_1 = -\frac{k}{\sqrt{w}} + \frac{\sqrt{w}}{2}.$$

The formula comes from Lemma 6.6: the price depends on $\sigma$ only through $w = \sigma^2 T$. So the chain rule gives $\mathcal{V} = e^{-rT} F_T \frac{\partial g}{\partial w} \frac{dw}{d\sigma} = e^{-rT} F_T \frac{\varphi(d_1)}{2\sqrt{w}} 2\sigma T = e^{-rT} F_T \sqrt{T}\, \varphi(d_1)$. The put has the same vega because put–call parity does not contain $\sigma$. The density $\varphi$ is largest at $0$ and symmetric. So the vega is largest at $k = w/2$, just above the money. It is the same at $k = w/2 + u$ and $k = w/2 - u$. It falls to zero in both wings.

For a market quote, $\sigma$ is the implied volatility $\sigma_{\text{imp}}(k, T)$ of that quote. Let the surface change by a small $\Delta\sigma_{\text{imp}}(k, T)$. To first order, the price of a call or a put at $(k, T)$ changes by $\mathcal{V}(k)\, \Delta\sigma_{\text{imp}}(k, T)$. So the vega tells how much a position earns when one point of the surface moves (Figure 7, left). For example, take 30 days and $\sigma = 20\%$. An option at $k = -0.05$ has 67% of the peak vega. An option at $k = 0.05$ has 70%.

**Level: the straddle.** The at-the-money straddle has the simplest price of the five positions.

**Proposition 6.9** (The at-the-money straddle).
*Let $\sigma = \sigma_{\text{imp}}(0, T)$ be the implied volatility at $k = 0$, and $w = \sigma^2 T$. At the strike $K = F_T$ the call and the put have the same price, and the straddle costs*

$$\Pi \;=\; 2e^{-rT} F_T \left( 2\Phi\left(\tfrac{1}{2}\sqrt{w}\right) - 1 \right) \;\approx\; e^{-rT} F_T \sqrt{\frac{2}{\pi}}\, \sigma\sqrt{T}.$$

The at-the-money straddle is a price for the level of the surface at $k = 0$. Its price is almost linear in implied volatility, with $\sqrt{2/\pi} \approx 0.798$. So one volatility point buys almost the same amount of straddle at every level. At 30 days and $\sigma = 20\%$, the straddle costs 4.57% of $e^{-rT} F_T$. The linear price is off by about 0.01%. Even at one year and $\sigma = 50\%$ it is off by only about 1%. Figure 7 (middle) shows the two prices. The straddle holds two options at $k = 0$, so its vega is $2\mathcal{V}(0)$. That is close to the largest vega at this maturity.

*Proof.*

□

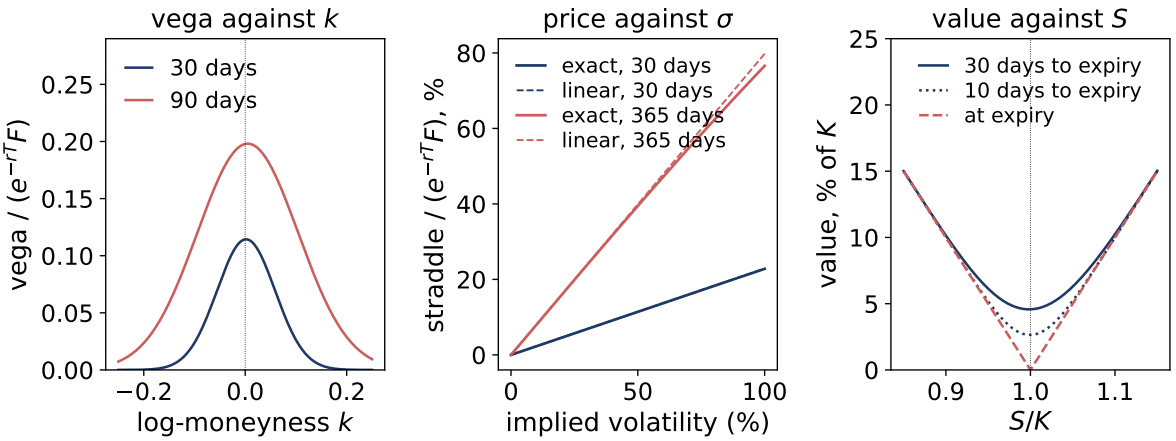



Figure 7: Left: the vega $\mathcal{V}(k)$ against $k$, at $\sigma = 20\%$, for 30 and 90 days. Middle: the at-the-money straddle price of Proposition 6.9 against implied volatility, exact (solid) and linear (dashed). At 30 days the two curves cannot be told apart. At one year they separate by 4% at $\sigma = 100\%$. Right: the value of a 30-day straddle at $K = F$ against $S/K$, at 30 days and 10 days to expiry and at expiry ($\sigma = 20\%$, $r = \delta = 0$). Near $K$ the value hardly depends on the direction of the move. Far from $K$ it moves almost one for one with $S$.

**Tails: the strangle and the butterfly.** The strangle pays only when $S_T$ ends below $K_1$ or above $K_2$. By risk-neutral pricing, its price uses only the two tails of the risk-neutral density $q$:

$$P(K_1) + C(K_2) = e^{-rT} \int_0^{K_1} (K_1 - s)\, q(s)\, ds + e^{-rT} \int_{K_2}^{\infty} (s - K_2)\, q(s)\, ds.$$

So the strangle is a position in the tails. Its vega sits at $k_1$ and $k_2$, the log-moneyness of its two strikes. So it gains when the wings of the smile rise.

The butterfly is a position in the density near $K$. Theorem 6.4 gives its price as about $e^{-rT} q(K) h^2$, and Proposition 6.7(a) says that price is never negative. Put–call parity links it to the two other positions. Take $K_1 = K - h$ and $K_2 = K + h$. Parity gives

$$C(K - h) = P(K - h) + e^{-rT}(F_T - K + h), \qquad C(K) = P(K) + e^{-rT}(F_T - K).$$

Put the first into the butterfly, and the second into one of its two $C(K)$:

$$C(K-h) - 2C(K) + C(K+h) = \big[P(K-h) + C(K+h)\big] - \big[C(K) + P(K)\big] + e^{-rT}h.$$

A long butterfly is therefore a long strangle, a short straddle and a bond that pays $h$. In options, it is long the wing options (the strangle) and short the center options (the straddle). In implied volatility, it is long the wings and short the at-the-money level. A butterfly at the money is short the level of volatility. Let $k_1$, $k$, $k_2$ be the log-moneyness of $K-h$, $K$, $K+h$. The butterfly's vega is $\mathcal{V}(k_1) - 2\mathcal{V}(k) + \mathcal{V}(k_2)$. Near the peak of the vega this is negative (Figure 7, left). For the strikes of Figure 6 it is $0.653 - 2(0.9996) + 0.713 = -0.633$ times the peak vega. So the butterfly gains when the wings of the smile rise relative to its center: when the smile becomes more curved.

**Skew: the risk reversal.** The long risk reversal holds the call at $k_2$ and is short the put at $k_1$. To first order, a change $\Delta\sigma_{\text{imp}}$ of the smile changes its price by

$$\mathcal{V}(k_2)\,\Delta\sigma_{\text{imp}}(k_2) - \mathcal{V}(k_1)\,\Delta\sigma_{\text{imp}}(k_1).$$

There are two cases. First, a parallel shift $\Delta\sigma_{\text{imp}}(k) = a$ changes the price by $\big(\mathcal{V}(k_2) - \mathcal{V}(k_1)\big)a$. This is zero when $k_1$ and $k_2$ lie at the same distance from $w/2$, because the vega is symmetric about $w/2$. Second, a tilt $\Delta\sigma_{\text{imp}}(k) = s\,k$, with $k_1 < 0 < k_2$, changes the price by

$$s\,\big(k_2\,\mathcal{V}(k_2) + |k_1|\,\mathcal{V}(k_1)\big),$$

which has the sign of $s$. So the risk reversal is a position in the skew, almost free of the level. On an index the skew is usually negative (fact 1 of §6.3). When it becomes more negative, the long risk reversal loses.

**Term structure: the calendar spread.** Let $\delta = 0$. Then $e^{-rT}F_T = S$ for every $T$. Write $g(k,w) = \Phi(d_1) - e^k\Phi(d_2)$, the normalized call price of Lemma 6.6. The calendar spread costs

$$C(K_2,T_2) - C(K_1,T_1) = S\,\Big[\,g\big(k, w(k,T_2)\big) - g\big(k, w(k,T_1)\big)\Big] \;\geq\; 0.$$

The sign follows from Proposition 6.7(b), because $w(k,T_2) \geq w(k,T_1)$ and $g$ is increasing in $w$ (Lemma 6.6). The price is zero exactly when $w(k,T_2) = w(k,T_1)$. The difference $w(k,T_2) - w(k,T_1)$ is the variance that the prices assign to the period from $T_1$ to $T_2$. So the calendar spread is a position in the term structure of $w$. At $k = 0$ the vega grows with maturity for every listed option ($w < 4$). So a long calendar spread at the money gains from a parallel rise in implied volatility. It loses when the short end rises much more than the long end. This happened on the crash day of Figure 2.

**The straddle and the variance risk premium.** Section 6.2 called the gap between implied volatility and the volatility later realized the variance risk premium. The straddle turns this gap into a return. The next result needs an assumption about the real-world distribution of $S_T$.

**Proposition 6.10** (Expected return of the straddle).
*Buy the straddle of Proposition 6.9 at time* 0*, with implied volatility* $\sigma_{\text{imp}} = \sigma_{\text{imp}}(0,T)$*, and hold it to* $T$*. Assume that under* $\mathbb{P}$

$$\ln \frac{S_T}{F_T} \;\sim\; N\Big(-\tfrac{1}{2}\sigma_R^2 T,\; \sigma_R^2 T\Big)$$

*for a constant* $\sigma_R > 0$*. Then the expected gross return of the straddle is*

$$R \;=\; \frac{\mathbb{E}^{\mathbb{P}}\big[\,|S_T - F_T|\,\big]}{\Pi} \;=\; e^{rT}\,\frac{2\Phi\big(\tfrac{1}{2}\sigma_R\sqrt{T}\big) - 1}{2\Phi\big(\tfrac{1}{2}\sigma_{\text{imp}}\sqrt{T}\big) - 1} \;\approx\; e^{rT}\,\frac{\sigma_R}{\sigma_{\text{imp}}}.$$

The assumption has two parts. First, $\sigma_R$ is the volatility that the index actually has under $\mathbb{P}$. The realized volatility of Definition 6.3 estimates it after the fact. Second, the mean is set so that $\mathbb{E}^{\mathbb{P}}[S_T] = F_T$. So the index earns no risk premium over $[0,T]$. The result then says that the straddle's gross return is about $e^{rT}\sigma_R/\sigma_{\text{imp}}$. The map $x \mapsto 2\Phi(x)-1$ is increasing. So the expected excess return is negative exactly when implied volatility exceeds the volatility the index delivers. For example, take $\sigma_{\text{imp}} = 20\%$, $\sigma_R = 16\%$, 30 days and $r = 0$. Then $R = 0.800$, so the straddle loses 20.0% on average. The no-premium part of the assumption matters little over 30 days. In this example, an expected excess return of 6% a year raises $\mathbb{E}^{\mathbb{P}}|S_T - F_T|$ by only 0.8%.

*Proof.*

□

So, under the assumption, the expected excess return of the straddle has the sign of $\sigma_R - \sigma_{\text{imp}}$. A positive variance risk premium makes the straddle lose on average. This is why straddle returns are used as a traded proxy for the premium. Coval and Shumway (2001) study straddles on the S&P 500 index near the money. Their average returns are large and negative. The authors read this as evidence that the market prices volatility risk.

One straddle return says little, however. Figure 8 simulates the trade under the assumption of Proposition 6.10. The mean is $-20.0\%$, as the proposition gives. But the median is $-32.2\%$, and 68.1% of the trades lose money. A few large moves pay several times the price, so the distribution is skewed to the right. Even with no premium at all, most straddles lose: with $\sigma_R = \sigma_{\text{imp}} = 20\%$, a trade loses with probability 57.5%. So the premium shows only in an average over many periods.

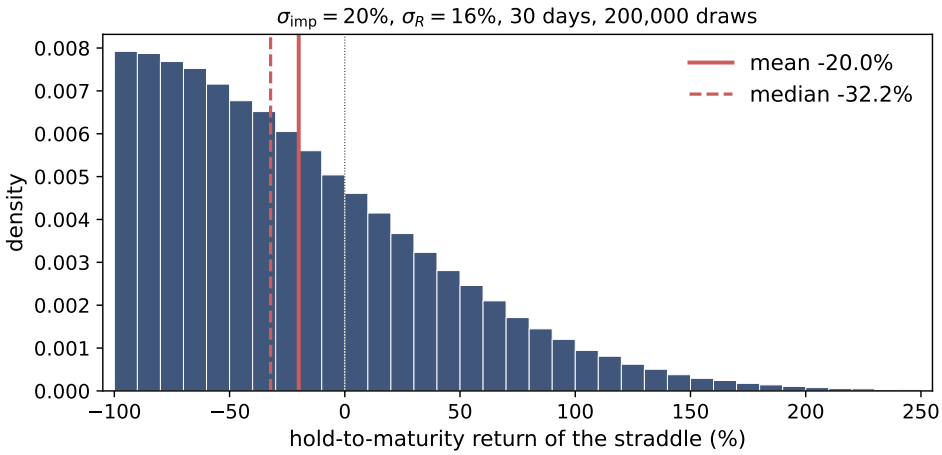



Figure 8: Simulated hold-to-maturity returns of a 30-day straddle at $K = F$, bought at $\sigma_{\text{imp}} = 20\%$ when the index has $\sigma_R = 16\%$. Log returns are normal as in Proposition 6.10, $r = 0$, 200,000 draws, seed 20261108. The mean return is $-20.0\%$ and the median $-32.2\%$. 68.1% of the trades lose money. The skewness is 1.01. The axis stops at 250%; the largest return is 415%.

Figure 9 runs the trade on TXO data. There is one trade per monthly expiry: 276 trades, from January 2002 to December 2024. Each trade buys the call and the put at the listed strike nearest the parity forward. The forward is the median of $K + e^{rT}(C - P)$ over strikes within 2% of the index close, with $r = 1.5\%$. In 48 thin months, with fewer than three such strikes, the window is 6%. Every strike used has a call and a put settlement price of at least 1. It enters about 30 calendar days before expiry (24 to 36 days, median 30). It costs the two settlement prices. It pays $|\text{FSP} - K|$, where FSP is the final settlement price that TAIFEX sets at expiry. The trade is held to expiry.

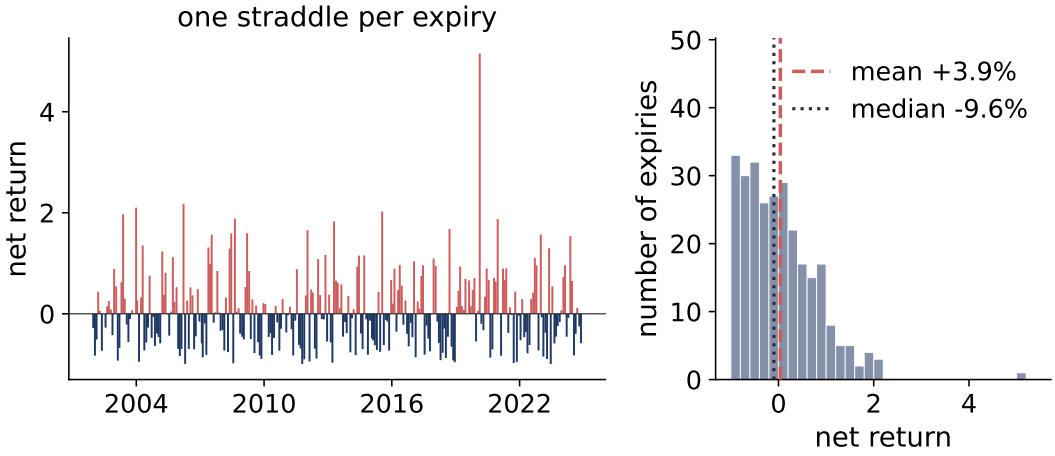



Figure 9: Net returns of 30-day TXO straddles at the strike nearest the parity forward. Each is held to expiry, one per monthly expiry, from January 2002 to December 2024 (276 trades). Left: each trade's return by date. Right: the distribution. The mean is $+3.9\%$ and the median $-9.6\%$. 53.6% of the trades lose money. The largest return is $+515\%$, in March 2020. Data: TAIFEX settlement prices; `figures/make_vol_trading_figures.py`.

In volatility units the premium is clear. The implied volatility of the trades averages 19.7%. The realized volatility of the TAIEX over the same days averages 17.2%. Realized volatility was

below implied volatility in 76.8% of the months. The mean of $\sigma_R/\sigma_{\text{imp}} - 1$ is $-9.6\%$, and its median is $-15.9\%$. Under the assumptions of Proposition 6.10, this gap gives an average net return of about $-10\%$.

The straddle returns do not show it clearly. The median trade loses 9.6%, and 53.6% of the trades lose money. But the mean net return is $+3.9\%$. The returns are very dispersed. Their standard deviation is 0.787, so the standard error of the mean is 0.047. The mean is only 0.8 standard errors from zero. So the data cannot tell it apart from zero. The largest return is $+515\%$, in March 2020. Without that one trade, the mean is $+2.0\%$ and the median is $-10.5\%$.

The link between one trade and its volatility gap is loose. Across trades, the correlation between the net return and $\sigma_R/\sigma_{\text{imp}} - 1$ is only 0.287. There are three reasons. First, the payoff depends only on the terminal move $|S_T - K|$. But $\sigma_R$ measures the daily moves along the path. A month with large daily swings can still end near the strike, and the straddle then loses. Second, index returns have fat tails, and the returns of the straddle are skewed (skewness 1.52 in the data). A rare large move pays several times the cost, as in March 2020. Such months move the sample mean much more than ordinary months. Third, there is only one draw per month. Even 23 years give only 276 trades, and Figure 8 shows how widely one draw varies. So the straddle return is a very noisy proxy for the premium. In the TXO data, it does not reproduce the negative mean that Coval and Shumway (2001) find for the S&P 500. The designs also differ. Coval and Shumway use zero-beta straddles and weekly returns. Here each straddle is held to expiry and bought at settlement prices, with no bid–ask cost.

Two more limits apply. First, Proposition 6.10 assumes normal log returns. With fat tails, $\mathbb{E}^{\mathbb{P}}|S_T - F_T|$ is no longer fixed by $\sigma_R$ alone. So the straddle prices a premium on the size of moves. This is close to the variance premium, but it is not the same. Second, the straddle is free of direction only at the start, with $K = F_T$. When $S$ moves away from $K$, its value moves almost one for one with $S$ (Figure 7, right). So a straddle closed before $T$ earns a mix of a volatility return and a directional return. The mix depends on the path of $S$. A trader can remove the directional part by trading the underlying against the straddle. That delta-hedged position is not covered in this section.

## 6.7 The VIX

This subsection explains the VIX, the volatility index of the Cboe (the Chicago Board Options Exchange). It first describes the original index of 1993 and its problems. It then states the formula used since 2003 and derives it. It ends with what the index means today and with Taiwan's version. Throughout, the interest rate $r$ and the dividend yield $\delta$ are constants. The forward price for maturity $T$ is $F = F_T = S_0 e^{(r-\delta)T}$.

**The old VIX.** The Cboe introduced its volatility index in 1993. Whaley built it for the exchange. Whaley (1993) proposed futures and options written on it. In 2003 the index was redefined, and the old index was renamed VXO. Carr and Wu (2006) describe the VXO as follows. It uses options on the S&P 100 index. At each of the two nearest maturities, take the two strikes closest to the spot, one above and one below. At each strike, average the Black–Scholes implied volatilities of the call and the put. Interpolate linearly between the two strikes to the spot. This gives an at-

the-money implied volatility $\mathrm{ATMV}(T_j)$ for $j = 1, 2$: eight options in all. If the nearest maturity is within eight calendar days, the next two maturities are used instead.

Next comes a "trading-day conversion". Let $N_C$ be the number of calendar days to expiry and $N_T$ the number of trading days. Then

$$\mathrm{TV}(T_j) = \mathrm{ATMV}(T_j)\sqrt{\frac{N_C}{N_T}}, \qquad N_T = N_C - 2 \times \mathrm{int}(N_C/7).$$

Finally, the two maturities are interpolated to 22 trading days:

$$\mathrm{VXO} = \mathrm{TV}(T_1)\,\frac{N_{T,2} - 22}{N_{T,2} - N_{T,1}} + \mathrm{TV}(T_2)\,\frac{22 - N_{T,1}}{N_{T,2} - N_{T,1}}.$$

A month has about 22 trading days. So the VXO is a one-month at-the-money implied volatility.

The VXO has four problems.

1. *It depends on a pricing model.* Each input is a Black–Scholes implied volatility. The Black–Scholes model assumes a constant volatility. So the model itself gives no reason to read the number as a forecast of future volatility.

2. *The trading-day conversion raises the level.* For a 30-day option, $N_C = 30$ and $N_T = 30 - 2\times 4 = 22$. So $\mathrm{TV} = \mathrm{ATMV}\sqrt{30/22} \approx 1.17\,\mathrm{ATMV}$. An at-the-money implied volatility of 20% enters the index as about 23.4%. The reason is a mix of conventions. The conversion spreads the total variance to expiry over the $N_T$ trading days. But it still annualizes with 365 days. So the VXO is not comparable with a realized volatility annualized in the usual way. Carr and Wu (2006) therefore rescale it by $\sqrt{22/30}$.

3. *It ignores the smile.* It uses only the strikes next to the spot. Section 6.3 showed that the out-of-the-money puts carry information of their own, about the left tail.

4. *No portfolio pays it.* An implied volatility is a nonlinear function of an option price. So the VXO is not the price of any fixed portfolio of options. A dealer who sells a contract on it cannot hedge that contract with a static position.

**The new VIX.** On 22 September 2003 the Cboe changed the definition. The new index uses S&P 500 options. It uses prices, not implied volatilities, at all available strikes. The Cboe methodology refers to Demeterfi, Derman, Kamal and Zou (1999) for the theory.

**Definition 6.11** (The VIX formula)**.**
*Fix a maturity $T$ and the listed strikes $K_1 < K_2 < \cdots < K_n$. Let $K_0$ be the first listed strike equal to or below the forward $F$. For each strike let*

$$Q(K_i) = \begin{cases} P(K_i), & K_i < K_0, \\ \frac{1}{2}\big(P(K_0) + C(K_0)\big), & K_i = K_0, \\ C(K_i), & K_i > K_0, \end{cases} \qquad \Delta K_i = \frac{K_{i+1} - K_{i-1}}{2},$$

*with $P$ and $C$ the mid quotes of puts and calls. At the lowest and the highest strike, $\Delta K_i$ is the distance to the one neighbor. The* single-maturity variance *is*

$$\sigma^2(T) = \frac{2}{T} \sum_i \frac{\Delta K_i}{K_i^2} e^{rT} Q(K_i) \; - \; \frac{1}{T}\left(\frac{F}{K_0} - 1\right)^2.$$

*Let $T_1 < T_2$ be two maturities, normally one on each side of $\tau = 30/365$, and put $w_1 = (T_2 - \tau)/(T_2 - T_1)$ and $w_2 = (\tau - T_1)/(T_2 - T_1)$. The* VIX *is*

$$\text{VIX} = 100 \times \sqrt{\frac{w_1 T_1 \sigma^2(T_1) + w_2 T_2 \sigma^2(T_2)}{\tau}}.$$

The letter $Q(K)$ is the Cboe's. It is an option price, not the measure $\boldsymbol{Q}$. The formula uses out-of-the-money options only: puts below $K_0$ and calls above it. Each price gets the weight $\Delta K_i / K_i^2$. The two maturities are joined linearly in the total variance $\sigma^2 T$, as the surface of §6.5 joins its maturities. Each maturity uses its own forward and its own rate; the formula suppresses this in the notation. The rest of the subsection shows what $\sigma^2(T)$ measures, and why each price gets the weight $1/K^2$.

**Proposition 6.12** (Model-free implied variance)**.**
*Under $\boldsymbol{Q}$ let the index follow*

$$\frac{dS_t}{S_t} = (r - \delta)\, dt + \sigma_t\, d\widetilde{W}_t,$$

*where the volatility $\sigma_t$ may be random. Let $C(K)$ and $P(K)$ be the prices of calls and puts with maturity $T$. Let $Q(K) = P(K)$ for $K < F$ and $Q(K) = C(K)$ for $K \geq F$. Then*

$$\mathbb{E}^{\boldsymbol{Q}}\left[\frac{1}{T}\int_0^T \sigma_t^2\, dt\right] = \frac{2e^{rT}}{T}\int_0^\infty \frac{Q(K)}{K^2}\, dK.$$

The left side is the risk-neutral expected variance per year over $[0, T]$. The right side uses only today's prices of out-of-the-money options. No model for $\sigma_t$ is needed. So option prices fix the expected variance, whatever the volatility process. The single-maturity variance of Definition 6.11 is a sum that approximates this integral. Britten-Jones and Neuberger (2000) show that option prices alone fix this expected variance when prices move continuously.

*Proof.*

□

Step 1 says that variance is a log contract. A *log contract* is a claim that pays $\ln(S_T/F)$ at $T$. Total variance equals twice the gain from holding $1/S_t$ units of the index, minus twice the log return. The first piece has a known risk-neutral mean. So the expected total variance is $-2e^{rT}$ times the price of a log contract. This holds for any volatility process, random or not.

Step 2 explains the weights $1/K^2$. The log payoff is a curve in $S_T$. A call or a put with strike $K$ has a payoff with one kink at $K$. There its slope changes by one. A smooth curve is built from many small kinks. At each strike we need as many options as the second derivative of the payoff at that strike. The second derivative of $\ln K$ is $-1/K^2$. So the log contract holds $1/K^2\,dK$ options at each strike $K$, sold short. Variance is minus twice the log contract, so it holds them long, twice over. The log curves most at low prices. So low strikes get the largest weight. Carr and Madan (1998) use the same idea, for any smooth payoff, to build volatility trades from options.

**From the integral to the formula.** The listed strikes rarely include $F$. So the Cboe splits puts and calls at $K_0$, the listed strike just below $F$. Between $K_0$ and $F$ it then uses calls that are in the money. The last term of Definition 6.11, $(F/K_0 - 1)^2/T$, is a small correction for this gap between $K_0$ and $F$. On $[K_0, F]$ the calls exceed the puts by $C - P = e^{-rT}(F - K)$. So the excess is $\frac{2}{T}\int_{K_0}^{F} \frac{F-K}{K^2}\,dK \approx \frac{1}{T}\left(\frac{F}{K_0} - 1\right)^2$. At $K_0$ itself the formula averages the put and the call. The reason is that the strip of width $\Delta K$ around $K_0$ lies half below $K_0$ and half above it.

So the VIX makes three approximations. First, the integral over all strikes becomes a sum over the listed strikes. The sum also stops at the lowest and the highest strike. Second, the $K_0$ term corrects the split at $K_0$ only to leading order. Third, the two maturities are joined linearly in total variance. This is exact when the expected variance rate is constant between $T_1$ and $T_2$.

**What the VIX means today.** The derivation gives the VIX a meaning as a price. A contract makes this precise.

**Definition 6.13** (Variance swap)**.**
*A* variance swap *with maturity $T$ and* variance strike $K_{\text{var}}$ *pays at $T$*

$$N \times \left(\text{RV}_{0,T} - K_{\text{var}}\right),$$

*where $N$ is a notional amount and $\text{RV}_{0,T}$ is the* realized variance *per year over $[0, T]$. In the model of Proposition 6.12, $\text{RV}_{0,T} = \frac{1}{T}\int_0^T \sigma_t^2\, dt$. The strike is set so that the contract costs nothing to enter.*

The buyer receives the variance that is realized and pays a fixed variance agreed today. So the swap is a bet on realized variance against the strike. In a contract, $\text{RV}_{0,T}$ is computed from daily returns. It is then the square of the realized volatility of Definition 6.3.

The fair strike follows in one line. The contract costs nothing, so its risk-neutral value is zero: $e^{-rT}\mathbb{E}^{Q}[N(\text{RV}_{0,T} - K_{\text{var}})] = 0$. So $K_{\text{var}} = \mathbb{E}^{Q}[\text{RV}_{0,T}]$, the expected variance. By Proposition 6.12, this is the strip of out-of-the-money options. So $(\text{VIX}/100)^2$ is the variance strike of a 30-day variance swap on the S&P 500, up to the three approximations above. The VIX is the square root of this strike, in percent. Carr and Wu (2006) give the same reading.

This reading needs continuous paths. When the index can jump down, $\text{VIX}^2$ lies slightly below the fair variance strike. The error is of third order in the jump size (Carr and Wu, 2006). Jiang and Tian (2005) study the same formula when prices can jump.

*What the weights mean.* In Proposition 6.12 each option price gets the weight $1/K^2$. A put at $K = 0.8F$ gets $1/0.64 \approx 1.56$ times the weight of an option at $F$. A call at $K = 1.2F$ gets $1/1.44 \approx 0.69$ times that weight. The smirk of §6.3 makes the low-strike puts expensive as well. So out-of-the-money puts carry a large weight in the VIX. These puts are insurance against a crash. So the VIX also prices crash protection. It is not only a forecast of variance. It can rise when puts become dearer, even if the at-the-money implied volatility does not move.

*The variance risk premium.* §6.2 named this gap. It can now be defined exactly.

**Definition 6.14** (Variance risk premium)**.**
*The* variance risk premium *at date $t$ over $[t, t+T]$ is*

$$\text{VRP}_t = \mathbb{E}^{Q}\left[\text{RV}_{t,t+T} \mid \mathcal{F}_t\right] - \mathbb{E}^{\mathbb{P}}\left[\text{RV}_{t,t+T} \mid \mathcal{F}_t\right].$$

*For $T = 30$ days the first term is $(\text{VIX}_t/100)^2$, up to the approximations above.*

The first term is the variance strike, the price of variance. The second is the variance that investors actually expect. So the premium is what a buyer of variance pays above the expected outcome. The seller of the swap earns it on average. The buyer pays it because variance tends to be high when the market falls. On the crash day of §6.3, the index fell and implied volatility jumped. A long variance position pays off in bad times, like insurance. The second term is not observed. So research measures the premium after the fact, as $(\text{VIX}_t/100)^2$ minus the realized variance that followed. Carr and Wu (2006) study the S&P 500 from 1990 to 2005. On average, the VIX was about 5 volatility points above the realized volatility of the next 30 days. Carr and Wu (2009) build variance swap rates from option prices for five stock indexes and 35 stocks. They write the premium as realized minus implied variance, with the opposite sign. They find it strongly negative for the stock indexes. Bollerslev, Tauchen and Zhou (2009) find that a high

premium predicts high stock returns over the next quarter. §6.6 used straddle returns as a proxy for this premium. A variance swap measures it directly, in units of variance.

**Taiwan's VIX.** The *TAIWAN VIX* is the volatility index of the Taiwan Futures Exchange (TAIFEX). It was launched on 18 December 2006. It applies the Cboe formula of Definition 6.11 to TXO. It uses the near-month and the next-month contracts. When the near month has two trading days left, the pair moves to the next two months. The forward $F$ is the price of the TX index future of the same month, not a value from put–call parity. $K_0$ is the highest strike at or below $F$. The option prices are mid quotes, after a filter for stale or extreme quotes. The two maturities are interpolated to 30 days, and the index is published every 15 seconds.

The top panel of Figure 3 shows the VIX itself, for the S&P 500. For TXO we use a related series for 2001–2024. It is `MF_Vol_30`, computed by the method of Bakshi, Kapadia and Madan (2003). It also integrates out-of-the-money TXO prices across strikes, with other weights. It estimates the standard deviation of the 30-day risk-neutral log return. The two series need not be equal, but both read the same option prices. Figure 10 compares `MF_Vol_30` with the volatility that the TAIEX realized over the next 30 days.

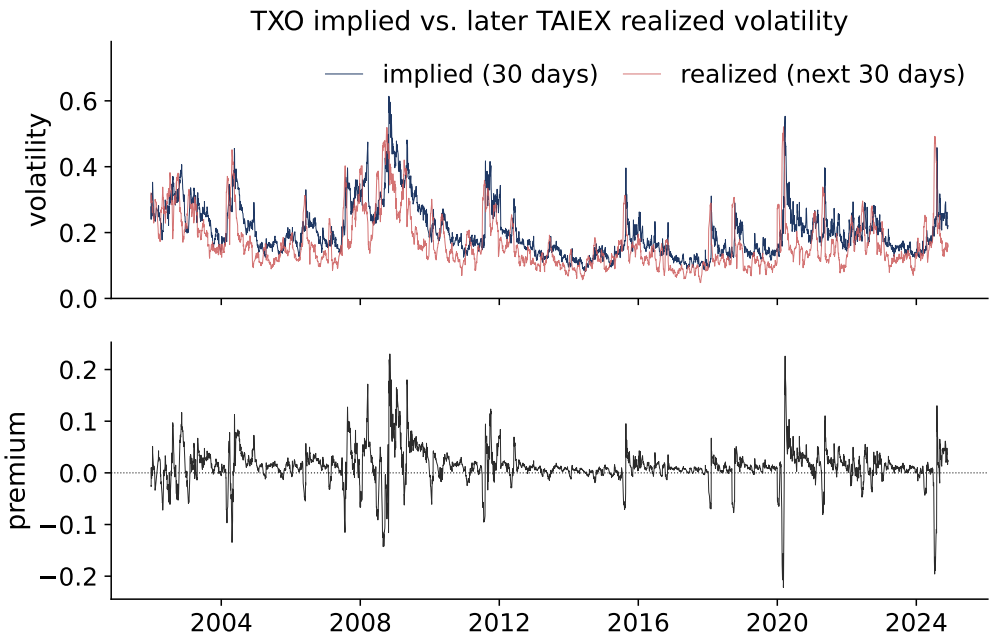



Figure 10: Top: TXO 30-day model-free implied volatility (`MF_Vol_30`, §6.7). It is compared with the realized volatility of the TAIEX over the next 30 calendar days. The sample has 5,019 days, from 2001-12-24 to 2024-11-29. The realized volatility applies Definition 6.3 to the daily log returns of TAIEX closes. Bottom: a proxy for the variance premium after the fact, $\texttt{MF_Vol_30}^2$ minus the realized variance that followed. The implied volatility averages 20.3% and the realized volatility 17.0%. The premium averages 0.0119 in variance per year (median 0.0107). It is positive on 79.3% of days. It is most negative on 2020-03-05: the implied volatility was 19.8%, and 51.1% was realized. It is most positive on 2008-11-07: 59.4% implied, 35.1% realized.

## 6.8 Local Volatility

This subsection builds the smallest change to Black–Scholes that can fit the whole surface. It shows that the surface fixes this model. It then shows why an exact fit is still not enough. In this subsection there are no dividends ($\delta = 0$).

**Definition 6.15** (Local volatility model)**.**
*In a* local volatility model *the stock follows, under* $\mathbb{Q}$*,*

$$dS_t = rS_t\,dt + \sigma_{\text{loc}}(t, S_t)\,S_t\,d\widetilde{W}_t,$$

*where* $\sigma_{\text{loc}}$ *is a deterministic function of time and the stock price.*

Volatility now changes with the stock price and with time, but it has no randomness of its own. There is still one source of risk. So the market is still complete, and every option can still be hedged with the stock alone. Dupire's result is that the surface fixes the function $\sigma_{\text{loc}}$.

**Proposition 6.16** (Dupire's formula)**.**
*In the model of Definition 6.15, let* $C(K,T)$ *be the call prices of all strikes and maturities. Assume that* $C$ *is smooth in* $(K,T)$ *and that* $\partial^2 C/\partial K^2 > 0$*. Then*

$$\sigma^2_{\text{loc}}(T,K) \;=\; \frac{\dfrac{\partial C}{\partial T} + rK\dfrac{\partial C}{\partial K}}{\dfrac{1}{2}K^2\dfrac{\partial^2 C}{\partial K^2}}.$$

The formula reads the local variance at the point $(T, K)$ off today's call prices. We do not prove it here. The proof needs the equation for how the density of $S_T$ changes with $T$ (the Fokker–Planck equation). It then needs several integrations by parts. The idea is simple. The density changes for two reasons: the drift carries probability along, and the diffusion spreads it out. Matching this change with the change of call prices in $T$ gives the formula (Dupire, 1994).

Each part of the formula is the price of a trade. The denominator is a butterfly. By Breeden–Litzenberger, $\partial^2 C/\partial K^2$ is the discounted density at $K$, and a butterfly spread prices it (§6.4). The numerator is a calendar spread. With $\delta = 0$, the strike $Ke^{r(u-T)}$ has the same log-moneyness $k$ at every maturity $u$. The numerator is the rate at which the call price grows along this line of fixed $k$:

$$\frac{\partial C}{\partial T} + rK\frac{\partial C}{\partial K} = \frac{d}{du}\,C\left(Ke^{r(u-T)}, u\right)\Big|_{u=T}.$$

So it is a calendar spread per unit of time. It is long a call at a later maturity and short a call at $T$, at the same moneyness. Proposition 6.7 signs both trades. The butterfly condition makes the denominator non-negative. The calendar condition makes the numerator non-negative, because a larger $w$ at the same $k$ means a larger price (Lemma 6.6). So an arbitrage-free surface gives a local variance that is not negative. A surface with static arbitrage can give a negative local variance, and no model has one.

So the surface of call prices fixes $\sigma_{\text{loc}}$: at most one local volatility model fits it. Dupire (1994) also shows the converse, which we do not prove either. If the surface is smooth and free of static arbitrage, the $\sigma_{\text{loc}}$ of the formula reproduces every quoted price. Exercise 6.10 checks the formula

on a flat surface.

An exact fit sounds like success. But the model makes a prediction that can be checked: how the smile moves when the stock moves. Since $\sigma_{\text{loc}}$ is a fixed function of $(t, S)$, the model has no freedom left about these moves. Hagan, Kumar, Lesniewski and Woodward (2002) show that the prediction has the wrong direction. When the underlying falls, the local volatility model moves the smile toward higher strikes. When it rises, the model moves the smile toward lower strikes. Observed smiles move in the same direction as the underlying. So local volatility prices today's vanilla options correctly. But its hedges depend on how the smile moves. So its hedges are wrong. It is an exact fit, but with the wrong dynamics. Section 6.9 returns to this point. Section 6.3 found one way to fit these data: a volatility that is a random process of its own. That needs a second source of risk. Section 7 adds it.

## 6.9 Using the Surface

This subsection shows four uses of the surface. Each one changes a number that a trading desk or a researcher actually computes.

**(A) Pricing off the surface.** A client may ask for an option that is not listed. For example, the client wants a 30-day option, but the nearest listed maturities are 16 and 51 days. The desk reads its volatility off the surface. At the option's moneyness it interpolates linearly in total variance $w$ between the two maturities (step 4 of §6.5). This keeps the calendar condition. Interpolating in $\sigma_{\text{imp}}$ instead gives a different number and has no such guarantee.

The level of the surface prices a vanilla option. Its slope prices a digital option. A *digital call* pays 1 if $S_T > K$ and 0 otherwise. Section 4 priced it at $e^{-rT}\Phi(d_2)$ under a flat smile. With a smile, the answer changes. In the rest of this subsection there are no dividends ($\delta = 0$).

**Proposition 6.17** (The digital call off the smile).
*Let $\delta = 0$ and fix a maturity $T$. Let the market call prices be $C(K) = C^{\text{BS}}\big(K, \sigma_{\text{imp}}(K)\big)$, with $\sigma_{\text{imp}}$ differentiable in $K$. Then the digital call with strike $K$ costs*

$$D(K) \;=\; e^{-rT}\,\Phi(d_2) \;-\; \mathcal{V}\,\frac{\partial \sigma_{\text{imp}}}{\partial K},$$

*where $d_2$ and the vega $\mathcal{V} = S\varphi(d_1)\sqrt{T}$ (Proposition 6.2) are evaluated at $\sigma_{\text{imp}}(K)$.*

The first term is the flat-smile price. The second term is the correction from the slope of the smile. On an index the slope is usually negative (fact 1), so the correction is positive. The digital call then costs more than the flat-smile price. The price of this exotic option depends on the slope of the surface, not only on its level. Exercise 6.6 computes both prices for one case.

*Proof.*

□

**(B) How the surface moves.** A surface describes today. A hedge also needs to know how the surface will move when the spot moves. Two simple rules are used in practice.

**Definition 6.18** (Sticky strike and sticky moneyness)**.**
*When the spot moves, the* sticky-strike *rule keeps the implied volatility of each strike $K$ unchanged. The* sticky-moneyness *rule keeps the implied volatility of each moneyness $K/S$ unchanged.*

Under sticky strike, the smile is fixed to the strikes, and the spot moves along it. Under sticky moneyness, the smile moves with the spot. Derman (1999) describes these rules and finds periods in which each one fits the data better. Sticky moneyness is also called sticky delta, because at a fixed maturity and volatility the delta depends only on moneyness.

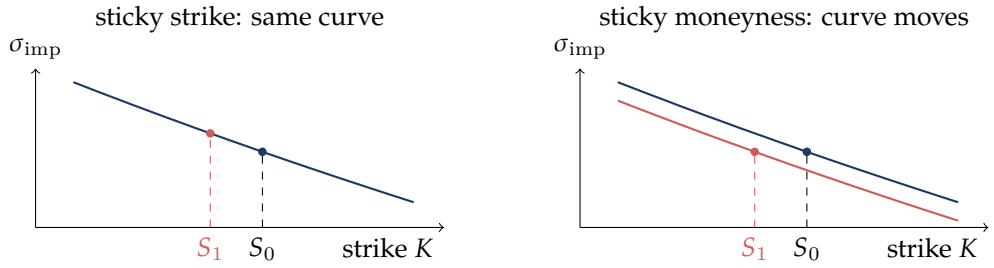



Figure 11: The spot falls from $S_0$ to $S_1$. Left: under sticky strike the smile stays put, so the new at-the-money volatility is read higher up the same curve. Right: under sticky moneyness the smile moves left with the spot, so the at-the-money volatility is unchanged. Exercise 6.4 puts numbers on both rules, for a 3% rise in the spot.

The local volatility model of §6.8 also implies a rule. Hagan et al. (2002) show that it moves the smile against the spot. Observed smiles move with the spot. So a desk that hedges with local volatility uses a rule that the data reject. Its vanilla prices still match the market.

**(C) Risk on the surface.** A book of options has an exposure to each part of the surface, not one exposure to "the" volatility. So desks split vega into *buckets* by maturity and by moneyness. Each

bucket uses the vega of Proposition 6.2, for the options in that bucket. A book can be flat to a parallel shift of the surface and still lose money when the term structure twists. The crash day of Figure 2 was such a twist: the short end rose far more than the long end.

**(D) The surface as a signal.** The surface also carries information about the underlying. Research uses it in five ways.

- *Crash risk.* The S&P 500 risk-neutral skewness is negative on every day of 2001–2024 (Figure 3). Bates (1991) finds that out-of-the-money index puts became unusually expensive in the year before the 1987 crash.
- *Stock returns.* Xing, Zhang and Zhao (2010) compare stocks by the steepness of their smirks. Stocks with the steepest smirks earn lower future returns than stocks with the flattest ones. Their reading is that informed traders with bad news buy out-of-the-money puts.
- *Option returns.* Expensiveness predicts option returns (Goyal and Saretto, 2009; §6.2).
- *Demand.* Bollen and Whaley (2004) find that net buying pressure changes the shape of the smile. So a steep skew need not reflect only beliefs about crashes.
- *Whole surfaces.* Kelly, Kuznetsov, Malamud and Xu (2023) use whole surfaces as inputs to machine learning models. They find information about future stock returns that the usual option measures miss.

**References.**

Bakshi, G., N. Kapadia and D. Madan (2003), *Review of Financial Studies* 16(1), 101–143.

Bates, D. S. (1991), *Journal of Finance* 46(3), 1009–1044.

Bollen, N. P. B. and R. E. Whaley (2004), *Journal of Finance* 59(2), 711–753.

Bollerslev, T., G. Tauchen and H. Zhou (2009), *Review of Financial Studies* 22(11), 4463–4492.

Breeden, D. T. and R. H. Litzenberger (1978), *Journal of Business* 51(4), 621–651.

Britten-Jones, M. and A. Neuberger (2000), *Journal of Finance* 55(2), 839–866.

Carr, P. and D. Madan (1998), Towards a theory of volatility trading, in R. Jarrow (ed.), *Volatility*, Risk Books.

Carr, P. and L. Wu (2006), A tale of two indices, *Journal of Derivatives*, Spring 2006, 13–29.

Carr, P. and L. Wu (2009), *Review of Financial Studies* 22(3), 1311–1341.

Cboe (2010), The CBOE Skew Index – SKEW, white paper.

Cboe (2026), Volatility Index Methodology: Cboe Volatility Index, version 6.0, 26 February 2026.

Christensen, B. J. and N. R. Prabhala (1998), *Journal of Financial Economics* 50(2), 125–150.

Coval, J. D. and T. Shumway (2001), Expected option returns, *Journal of Finance* 56(3), 983–1009.

Demeterfi, K., E. Derman, M. Kamal and J. Zou (1999), More than you ever wanted to know about volatility swaps, Goldman Sachs Quantitative Strategies Research Notes.

Derman, E. (1999), Regimes of volatility, Goldman Sachs Quantitative Strategies Research Notes.

Dupire, B. (1994), Pricing with a smile, *Risk*.

Gatheral, J. and A. Jacquier (2014), *Quantitative Finance* 14(1), 59–71.

Goyal, A. and A. Saretto (2009), *Journal of Financial Economics* 94(2), 310–326.

Hagan, P. S., D. Kumar, A. S. Lesniewski and D. E. Woodward (2002), Managing smile risk, *Wilmott Magazine*.

Jiang, G. J. and Y. S. Tian (2005), *Review of Financial Studies* 18(4), 1305–1342.

Kelly, B. T., B. Kuznetsov, S. Malamud and T. A. Xu (2023), Deep learning from implied volatility surfaces, Swiss Finance Institute Research Paper 23-60.

Taiwan Futures Exchange (TAIFEX), TAIWAN VIX: questions and answers (vixQA web page) and the methodology document VIX_book.pdf.

Whaley, R. E. (1993), Derivatives on market volatility: hedging tools long overdue, *Journal of Derivatives* 1, 71–84.

Xing, Y., X. Zhang and R. Zhao (2010), *Journal of Financial and Quantitative Analysis* 45(3), 641–662.

## Exercises

**Exercise 6.1.** *The last five daily TAIEX log returns are* $1\%$, $-2\%$, $0.5\%$, $1.5\%$ *and* $-1\%$. *Compute the realized volatility of Definition 6.3. The 30-day implied volatility is* $18\%$. *Is the option cheap or expensive by this measure? Why should you not trust the answer much?*

**Exercise 6.2.** *Sketch the vega* $S\varphi(d_1)\sqrt{\tau}$ *of Proposition 6.2 as a function of* $\ln(S/K)$. *Explain two things. (a) Where does a quoted price pin down the implied volatility best? (b) Why are deep out-of-the-money implied volatilities, the wings of Figure 1, fragile?*

**Exercise 6.3.** *Take a call and a put with the same strike and maturity. Use only put–call parity (Section 4) to show that they have the same implied volatility. Why does this not depend on any model?*

**Exercise 6.4.** *Today the smile at 51 days is* $\sigma_{\text{imp}}(K) = 0.18 - 0.40\,\ln(K/S_0)$ *with* $S_0 = 23000$. *The spot rises 3% to* $S_1 = 23690$. *Give the new at-the-money volatility under each rule of Definition 6.18.*

**Exercise 6.5.** *With* $r = 0$, *three calls are quoted:* $C(9600) = 428$, $C(9800) = 310$, $C(10000) = 200$. *(a) Estimate* $\boldsymbol{Q}(S_T > 9900)$ *and the density at* $9800$. *(b) State the convexity condition these three prices must satisfy, and check it.*

**Exercise 6.6.** *Take* $\sigma_{\text{imp}} = 18\%$ *at the money,* $\tau = 51/365$, $r = 0$, *no dividends and* $K = S = F$. *The slope in log-moneyness is* $\partial\sigma_{\text{imp}}/\partial k = -0.40$. *(a) Use Proposition 6.17 to price the at-the-money digital call with a flat smile and with this slope. (b) A digital put pays* $1$ *if* $S_T \le K$. *Explain why a digital call and a digital put with the same strike are together worth* $e^{-rT}$. *Price the digital put both ways.*

**Exercise 6.7.** *Figure 1 gives 22.1% at moneyness 0.90 on the calm day and 40.5% on the crash day. The at-the-money values of Figure 2 are 17.8% and 39.5%. Compute the slope* $(\sigma_{90} - \sigma_{100})/0.10$ *on each day. Both figures use the parity forward. Let S be the index close. With other choices of the forward the crash value changes. It is* $0.44$ *at* $F = S$. *For* $F/S$ *between 0.970 and 0.989, the range of that day's parity forwards, it runs from* $-0.14$ *to* $0.39$. *The calm value stays between* $0.36$ *and* $0.43$. *This holds at* $F = S$ *and for every* $F/S$ *in that day's range, 0.997 to 1.000. What can you conclude on each day?*

**Exercise 6.8.** *Let* $w_1 = w(k, T_1)$ *and* $w_2 = w(k, T_2)$ *with* $T_1 < T_2$. *(a) Show that the forward volatility* $\sqrt{(w_2 - w_1)/(T_2 - T_1)}$ *is a real number exactly when the calendar condition holds between* $T_1$ *and* $T_2$. *(b) Show that linear interpolation in w between* $T_1$ *and* $T_2$ *keeps the calendar condition. (c) On the crash day Figure 2 gives* $58.0\%$ *at 16 days and* $39.5\%$ *at 44 days. Compute the forward volatility between them. Compare it with the two implied volatilities. What does the market expect for volatility after the first 16 days?*

**Exercise 6.9.** *Check crash-day data quality at strike* $19200$, *with* $r = 0$ *and no dividends (so one strike has the same moneyness at both maturities). (a) At* $r = 0$ *the call/put-mean implied volatilities are* $41.4\%$ *at 44 days and* $27.9\%$ *at 72 days. (These are not Figure 2's parity-forward values.) Compute w at both maturities. Does the calendar condition look broken? (b) At* $r = 0$ *the raw calls (1200 at 44 days, 1240 at 72 days) have implied volatilities* $44.4\%$ *and* $35.9\%$. *The raw puts (1000 and 660) have* $38.3\%$ *and* $19.9\%$. *Compute w from the calls alone and from the puts alone. Which side breaks the calendar condition? Check the same answer directly in prices. (c) The total open interest is 8878 contracts at 44 days and 1320 at 72 days. At 72 days the put at 19200 has open interest 30, and the call has open interest 0. Why is this not proof of free money?*

**Exercise 6.10.** *Verify Dupire's formula in the flat world. Let $C(K,T)$ be the Black–Scholes price with a constant $\sigma$ for every $(K,T)$. Show that $\sigma^2_{\mathrm{loc}}(T,K) = \sigma^2$ everywhere. (Use the maturity derivative from Section 4's Greeks, $\partial C/\partial T = S\varphi(d_1)\sigma/(2\sqrt{T}) + rKe^{-rT}\Phi(d_2)$, and Step 3 of the proof of Proposition 6.17.)*

**Exercise 6.11.** *Take $r = 0$ and suppose the implied volatility depends on maturity only, $\sigma_{\mathrm{imp}}(T)$. Put $W(T) = \sigma^2_{\mathrm{imp}}(T)\,T$, which is non-decreasing by the calendar condition. Show that Dupire's formula gives $\sigma^2_{\mathrm{loc}}(T) = W'(T)$, the same at every strike. Then show that the forward volatility between two maturities is the root-mean-square local volatility over that period.*

**Exercise 6.12.** *Differentiate the Black–Scholes price twice with respect to $K$. Show that Breeden–Litzenberger then returns exactly the lognormal density of Section 3's geometric Brownian motion. Conclude that a flat smile goes with lognormal risk-neutral beliefs, and only with them.*