Bocconi Students Investment Club

Implied Volatility Surface Modelling

A Quantitative Framework for Construction, Calibration and Risk Management

Editors

Bocconi Students Investment Club

Paride Lauretti (Project Lead)

Riccardo Favazza

December 2025

bsic.it — Bocconi University, Milan

Abstract

This project develops a coherent framework for modelling the implied volatility surface by combining stochastic volatility, local volatility, and smile-consistent interpolation techniques. Starting from market conventions, risk-neutral valuation, and vanilla option pricing, the work introduces the volatility surface as a mapping from strikes and maturities to implied volatilities and formalizes the no-arbitrage constraints it must satisfy. The Heston model is then used to generate parametric volatility surfaces and to illustrate how each parameter affects skew, smile, and term structure, while the SVI family of parameterizations provides a flexible, arbitrage-aware representation of implied total variance across log-moneyness. In parallel, the Dupire local volatility framework links the implied surface to a state- and time-dependent instantaneous volatility, and the Vanna–Volga methodology is presented as a practical, market-driven correction to Black–Scholes prices that reproduces smile features from a small set of liquid quotes. Together, these tools yield a practical calibration and construction pipeline that can fit observed volatility data across asset classes, enforce static arbitrage conditions, and generate robust surfaces suitable for pricing and hedging both vanilla and first-generation exotic options.

Contents

1 Preliminaries

1.1 Notation and Conventions

In options and volatility surface modeling, it is essential to have a clear understanding of standard notation and market conventions. The spot price, denoted as S_0, refers to the current market value of the underlying asset. The strike price, often written as K, is the agreed-upon price at which the option can be exercised. Expiry, or maturity, designates the specific date when the option contract terminates, and at which payoff is determined. Time to maturity is sometimes denoted by T, calculated as the duration between the current time and expiry.

The forward price, F, is the contractually agreed future price of the asset, factoring in spot price, interest rates, and any dividends. The risk-free rate appears as r, used in discounted cash flow calculations and in models like Black–Scholes. Implied volatility, denoted as σ_{IV}, is a critical measure derived by inverting the Black–Scholes formula using observed market prices for options, settling on the volatility that best fits market pricing.

Options are classified primarily into calls and puts. A call grants the holder the right to buy the asset at strike, while a put grants the right to sell. These basic definitions set the groundwork for pricing models and for understanding the construction of the implied volatility surface, which is defined as a mapping between all strikes and expiries and their corresponding implied volatilities. Accordingly, quoting conventions differ across markets: equity surfaces are commonly quoted by strike, while FX markets often quote by option delta, impacting data interpretation and modeling approaches.

The main purpose of the paper is to illustrate the theoretical foundations of these techniques, and, in addition, we implement them on SPX data from June 2023 retrieved from WRDS as an empirical application.

1.2 Probability and Stochastic Calculus

The basics of stochastic calculus are at the heart of modeling asset price dynamics and deriving option prices. A standard approach is to model the underlying asset using geometric Brownian motion:

$$dS_t = \mu S_t\,dt + \sigma S_t\,dW_t$$

where S_t is the asset price, μ its drift, σ its volatility, and W_t is a standard Brownian motion representing the randomness in the price evolution. For a function $G(S, t)$ of the price and time, Itô's lemma gives its stochastic evolution:

$$dG = \frac{\partial G}{\partial t}dt + \frac{\partial G}{\partial S}\,dS + \frac{1}{2}\frac{\partial^2 G}{\partial S^2}\sigma^2 S^2\,dt$$

This result is essential for deriving the Black–Scholes PDE and for working with processes beyond simple Brownian motion.

No-arbitrage pricing in these models often uses a risk-neutral measure, where expected returns are discounted at the risk-free rate rather than the actual drift μ. Martingales play a critical role; under the risk-neutral measure, the discounted asset price must be a martingale:

$$E^Q\left[\frac{S_T}{P(0,T)}\middle|\mathcal{F}_t\right] = S_t$$

where Q denotes the risk-neutral measure and $P(0,T)$ is the discount factor for time T.

A general Itô process can be written as

$$dx_t = a(x_t,t)\,dt + b(x_t,t)\,dW_t$$

where the drift $a(x_t,t)$ and volatility $b(x_t,t)$ can depend on the current state and time.

Well-posedness requires Lipschitz conditions and suitable growth bounds on these functions to guarantee a solution exists. For option pricing, these mathematical tools allow one to transition from modeling the asset price to expressing the price of derivatives as solutions of partial differential equations or expected values under the risk-neutral measure.

1.3 Fundamentals of Risk-Neutral Pricing

Modern option pricing is grounded in the idea that, in an arbitrage-free market, the value of a derivative is determined by its expected payoff discounted at the risk-free rate, calculated under the so-called risk-neutral measure. This approach guarantees consistency with market prices and the principle of no-arbitrage, forming the foundation of both analytical formulas and numerical models.

In practice, let us consider a European option with maturity T and payoff H_T (for example, a call option has $H_T = \max(S_T - K, 0)$), with the present time denoted as $t \leq T$. If r is the continuously compounded risk-free rate, the fair value of the option at time t is given by:

$$C_t = e^{-r(T-t)} E^Q[H_T \mid \mathcal{F}_t]$$

$E^Q[\cdot]$ denotes expectation under the risk-neutral measure Q, and $\mathcal{F}_t$ is the information available up to time t. The discount factor $e^{-r(T-t)}$ ensures the payoff is fairly valued in today's terms.

The principle can be interpreted in two tightly linked ways:

- As a risk-neutral expectation (no-arbitrage pricing) of future payoffs, discounted at the risk-free rate.
- As the unique solution to a partial differential equation (like the Black–Scholes PDE) satisfied by the option price, with the terminal condition matching the option's payoff.

The famous Black–Scholes formula arises directly from these ideas, assuming the asset follows a geometric Brownian motion and there are no arbitrage opportunities, the price of a European option is fully determined by its expected discounted payoff under risk-neutral probabilities. This framework also underpins binomial tree models and Monte Carlo simulation approaches, where the expected value of the payoff is computed by averaging over all risk-neutral scenarios and then discounting back to present value.

1.4 Vanilla Options and Implied Volatility

Vanilla options, mainly European calls and puts, are the starting point for understanding implied volatility and its surface. Under the Black–Scholes model, these options can be priced by a closed-form formula, meaning you can directly compute the fair option price using market data and the model's parameters, rather than relying on simulations or numerical methods.

The key idea is that, given the input parameters (spot price S, strike K, risk-free rate r, time to expiry T and volatility σ), the price of a European call is given by the Black–Scholes formula:

$$C = SN(d_1) - Ke^{-rT}N(d_2)$$

where:

$$d_1 = \frac{\ln(S/K) + \left(r + \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}, \qquad d_2 = d_1 - \sigma\sqrt{T}$$

Here, $N(\cdot)$ is the cumulative distribution function of the standard normal distribution.

Implied volatility σ_{IV} is the unique volatility value that, when plugged into the Black–Scholes formula, makes the modeled price equal to the observed market price of the option. Market practitioners observe actual option prices and use numerical methods to solve for σ_{IV}, as it usually cannot be "solved for" in closed form. The result is a mapping across strikes and expiries, called the implied volatility surface.

The volatility surface displays how implied volatility changes with strike and time to expiry. This is crucial for practitioners: even though Black–Scholes assumes constant volatility, markets consistently show "smiles" and "skews", meaning implied volatility varies across different strikes and maturities.

1.5 Static No-Arbitrage Conditions (with Breeden–Litzenberger Theorem)

To ensure European option prices (and the resulting implied volatility surface) are free of static arbitrage, several key mathematical conditions must be satisfied:

1. **Positivity and Monotonicity**

   - Option prices must be non-negative: for all strikes and maturities, the price of a call option must satisfy $C(K, T) \geq 0$.
   - Monotonicity in strike: as the strike increases, the price of a call decreases:

   $$\frac{\partial C(K, T)}{\partial K} \leq 0.$$

2. **Convexity in Strike**

   Convexity: the price of a call option as a function of strike must be convex:

   $$\frac{\partial^2 C(K, T)}{\partial K^2} \geq 0.$$

   This has a probabilistic meaning via the Breeden–Litzenberger theorem:

The second derivative of the European call price with respect to strike is the risk-neutral probability density at strike

$$\frac{\partial^2 C(K,T)}{\partial K^2} = p_T(K)$$

Here, $p_T(K)$ is the risk-neutral probability density for the underlying asset at maturity T and strike K. Hence, convexity ensures this density is non-negative, which is necessary for the absence of arbitrage.

3. **Monotonicity in Expiry**

   Option prices must not decrease with expiry: for fixed strike, increasing expiry should not decrease the price:

$$C(K,T_2) \geq C(K,T_1) \quad \text{for } T_2 > T_1.$$

   This prevents "calendar spread" arbitrage, where a longer-dated option would be somehow cheaper than a shorter-dated one.

1.6 Smile Dynamics & Quoting Conventions

1. Smile and Skew Dynamics Beyond static arbitrage, a crucial aspect of implied volatility surfaces is their smile and skew dynamics, how volatility varies as you move across strikes or as market conditions change. The "smile" refers to cases where implied volatility is highest for deep in-the-money or out-of-the-money options, while the "skew" describes asymmetric changes, such as implied volatility increasing steadily for out-of-the-money puts or calls.

2. Sticky-strike vs Sticky-delta A central issue is how the surface reshapes when the underlying asset moves.

- In sticky-strike quoting, the surface is fixed with respect to strike; as spot moves, the implied volatility for a given strike remains the same.
- In sticky-delta quoting (common in FX), the surface is anchored by delta; as spot changes, the option moves along the surface so its delta remains constant.

3. Dividend and Foreign Rate Adjustments For assets with dividends or in cross-currency markets, forward price and discounting conventions affect delta definitions and surface shape. Practitioners must carefully map market-quoted volatilities onto their modeling surface, adjusting for these conventions.

4. Term Structure Intuition Over different expiries, volatility surfaces typically show a "term structure", for example with short-term volatility reflecting urgent news and longer-term volatility anchored to historic averages or implied market uncertainty.

1.7 The Volatility Surface Problem

Observed option prices for a range of strikes and expiries yield a grid of implied volatilities. Instead of a single value, we get a surface:

$$\sigma_{\text{IV}}(K, T)$$

where implied volatility is a function of both strike and time to expiry. This empirical surface reflects the risk preferences and price formation mechanisms of the actual market, which fundamentally differ from the lognormal assumptions of Black–Scholes. Smiles and skews, which are persistent features in equities, interest rates, FX, and commodities, signal these discrepancies.

The challenge, then, is to construct a *continuous, arbitrage-free implied volatility surface* from discrete option quotes. This surface must:

- Capture observed smiles/skews and term structure patterns.
- Enable pricing for options at intermediate or extrapolated strikes/maturities.
- Respect no-arbitrage conditions in both strikes and expiry.

Solution Approaches Constructing this surface is itself an extensive modelling task. Techniques fall broadly into two categories:

1. **Models for the underlying dynamics:** local volatility models, stochastic volatility structures (like Heston and SABR), and jump-diffusion processes calibrate parameters so their theoretical prices best fit the observed market IV grid.
2. **Direct surface modeling:** parametric, semi-parametric, or non-parametric interpolation schemes (SVI, splines, etc.) fit a smooth, arbitrage-free surface to the market data.

Both approaches must deal with practical concerns:

- Enforcing arbitrage-free conditions in strike and time.
- Robust extrapolation outside the core region.
- Calibration algorithms and fitting error control.

The resulting volatility surface becomes the backbone for pricing, risk management, and hedging, used by traders and risk managers not only for vanilla options but as a key input for exotics, portfolio risk analysis, and scenario testing.

2 Heston

2.1 Introduction: Heston Model

The **Heston Model** (1993) represents one of the most important extensions of the Black–Scholes framework.

While the Black–Scholes model assumes that asset returns are normally distributed with **constant volatility**, empirical evidence shows that this assumption fails to capture key features observed in financial markets, notably the **volatility smile** and **skewness** of implied volatilities.

Heston proposed a model where **volatility itself is stochastic**, meaning it evolves randomly over time instead of remaining fixed.

The model introduces a **correlation** between the asset's price and its volatility, allowing it to reproduce the asymmetry of option prices across different strikes.

From Constant to Stochastic Volatility

- The Black–Scholes model performs poorly for assets such as currencies or equities during volatile periods.
- Real-world volatility is not constant but **time-varying and mean-reverting**.
- Existing stochastic volatility models (Hull & White, Wiggins, Scott) lacked a **closed-form solution**, making them computationally heavy.
- Heston's approach provides a **semi-closed analytical solution** using **characteristic functions**, allowing for **efficient calibration** to market data.

This framework directly connects the behavior of volatility to the observed deviations in market option prices, specifically, how option values diverge from Black–Scholes predictions across different strikes and maturities.

Moreover, the model's structure is versatile, allowing it to be naturally extended to applications such as **bond** and **foreign currency options**, demonstrating its broad analytical flexibility.

2.2 Theory behind

The Heston model builds on the idea that volatility is not constant but **stochastic**, meaning it changes randomly over time.

It captures two essential aspects of financial markets:

1. **Mean reversion:** volatility tends to revert to a long-term average.
2. **Correlation:** volatility and asset returns are often negatively correlated.

2.2.1 Stochastic Processes

The model assumes that the asset price S_t and its variance v_t follow two coupled stochastic differential equations:

$$dS_t = \mu S_t\,dt + \sqrt{v_t}\,S_t\,dW_{1t}$$
$$dv_t = \kappa(\theta - v_t)\,dt + \sigma\sqrt{v_t}\,dW_{2t}$$
$$dW_{1t}\,dW_{2t} = \rho\,dt$$

where:

Parameter	Meaning
κ	Speed of mean reversion
θ	Long-term mean of variance
σ	Volatility of volatility
v_0	Initial variance
ρ	Correlation between price and volatility shocks

This formulation ensures that v_t remains positive (Cox–Ingersoll–Ross process) and fluctuates around its long-run average θ.

2.2.2 Risk-Neutral Valuation

Under the **risk-neutral measure**, the drift of the asset becomes $r - q$ (risk-free rate minus dividend yield), and a **volatility risk premium** λ is introduced:

$$dv_t = \kappa(\theta - v_t)\,dt + \sigma\sqrt{v_t}\,dW_{2t}^{Q}$$

$$dW_{1t}\,dW_{2t}^{Q} = \rho\,dt$$

The adjusted ("risk-neutral") parameters are:

$$\kappa^* = \kappa + \lambda, \quad \theta^* = \frac{\kappa\theta}{\kappa + \lambda}$$

These determine how volatility evolves under the pricing measure, influencing option values rather than actual market paths.

2.2.3 Pricing Framework

Any derivative $U(S, v, t)$ must satisfy the following **partial differential equation** derived from no-arbitrage arguments (Black–Scholes, Merton): Solving this PDE directly is complex, but Heston derived a **semi-closed form** solution using the **characteristic function** of the log-price process — a breakthrough that made stochastic volatility models practical for calibration

$$\frac{1}{2}vS^2\frac{\partial^2 U}{\partial S^2} + \rho\sigma vS\frac{\partial^2 U}{\partial S\partial v} + \frac{1}{2}\sigma^2 v\frac{\partial^2 U}{\partial v^2} + rS\frac{\partial U}{\partial S} + \left[\kappa(\theta - v) - \lambda(S, v, t)\right]\frac{\partial U}{\partial v} - rU + \frac{\partial U}{\partial t} = 0$$

2.2.4 European Call Option Pricing

A European call option with strike price K and maturity T satisfies the pricing PDE derived earlier, subject to the following boundary conditions:

$$U(S, v, T) = \max(0, S - K), \quad U(0, v, t) = 0,$$

and

$$\frac{\partial U}{\partial S}(\infty, v, t) = 1, \quad U(S, 0, t) = 0, \quad U(S, \infty, t) = S.$$

Solution Form By analogy with the Black–Scholes model, Heston proposed that the option price can be expressed as:

$$C(S, v, t) = SP_1 - Ke^{-r(T-t)}P_2$$

where the first term represents the present value of the underlying asset (weighted by P_1), and the second term represents the discounted strike price (weighted by P_2).

Both P_1 and P_2 are **risk-adjusted probabilities** under two different measures.

It is convenient to rewrite the equations in terms of the logarithm of the asset price:

$$x = \ln(S)$$

Partial Differential Equations for P_1 and P_2

Substituting the assumed solution into the general PDE leads to two coupled PDEs, one for each P_j with $j = 1, 2$

$$\frac{1}{2}v\frac{\partial^2 P_j}{\partial x^2} + \rho\sigma v\frac{\partial^2 P_j}{\partial x \partial v} + \frac{1}{2}\sigma^2 v\frac{\partial^2 P_j}{\partial v^2} + (r + u_j v)\frac{\partial P_j}{\partial x} + (a_j - b_j v)\frac{\partial P_j}{\partial v} + \frac{\partial P_j}{\partial t} = 0$$

where:

$$u_1 = \frac{1}{2}, \quad u_2 = -\frac{1}{2}$$

and the parameters are defined as:

$$a = \kappa\theta, \quad b_1 = \kappa + \lambda - \rho\sigma, \quad b_2 = \kappa + \lambda$$

The terminal condition for each probability function is given by:

$$P_j(x, v, T; \ln K) = \mathbf{1}_{\{x \geq \ln K\}}$$

2.2.5 Characteristic Function Solution

To obtain a closed-form expression for the option price, Heston reformulates the model in terms of the **log of the asset price**, $x = \ln(S)$.

Under the risk-neutral measure, the dynamics of x_t and v_t are given by:

$$dx_t = [r + u_j v_t]\,dt + \sqrt{v_t}\,dz_1(t)$$

$$dv_t = [a_j - b_j v_t]\,dt + \sigma\sqrt{v_t}\,dz_2(t)$$

where the parameters u_j, a_j, b_j are defined as before, and dz_1, dz_2 are correlated Wiener processes with correlation ρ.

The conditional probability that the option expires in the money is:

$$P_j(x, v, T; \ln K) = \Pr[x(T) \geq \ln K \mid x(t) = x, v(t) = v]$$

Characteristic Function Representation Although these probabilities are not available in closed form, they can be obtained through the **characteristic function** $f_j(x, v, T; \phi)$, which satisfies the same PDE as P_j but under Fourier transformation.

The characteristic function solution is expressed as:

$$f_j(x, v, t; \phi) = \exp\big(C_j(\tau; \phi) + D_j(\tau; \phi)v + i\phi x\big)$$

with:

$$C_j(\tau; \phi) = ri\phi\tau + \frac{a}{\sigma^2}\Big[(b_j - \rho\sigma i\phi + d)\tau - 2\ln\Big(\frac{1 - ge^{d\tau}}{1 - g}\Big)\Big]$$

$$D_j(\tau; \phi) = \frac{b_j - \rho\sigma i\phi + d}{\sigma^2}\Big(\frac{1 - e^{d\tau}}{1 - ge^{d\tau}}\Big)$$

and

$$g = \frac{b_j - \rho\sigma i\phi + d}{b_j - \rho\sigma i\phi - d},$$

$$d = \sqrt{(\rho\sigma i\phi - b_j)^2 - \sigma^2(2u_j i\phi - \phi^2)}.$$

Recovering Risk-Neutral Probabilities Once the characteristic function is known, the risk-neutral probabilities P_1 and P_2 are obtained by **Fourier inversion**:

$$P_j(x, v, T; \ln K) = \tfrac{1}{2} + \frac{1}{\pi}\int_0^\infty \Re\Big[\frac{e^{-i\phi\ln K} f_j(x, v, T; \phi)}{i\phi}\Big]\,d\phi$$

The integrand is a smooth, rapidly decaying function that can be efficiently evaluated numerically. The equations before together provide the **semi-closed form solution** for European call options under stochastic volatility.

2.2.6 From Heston Model to the Volatility Surface

The closed-form Heston formula allows us to compute European option prices

$$C(S_0, v_0, 0; K, T)$$

for any strike K and maturity T.

However, markets quote **implied volatilities**, so the connection to the volatility surface comes from **Black–Scholes inversion**. Therefore, the bridge between the Heston model and the **volatility surface** is the concept of *implied volatility*.

Implied Volatility Given a Heston price $C_{Heston}(K,T)$, the **Black–Scholes implied volatility** is defined as the value $\sigma_{imp}(K,T)$ such that:

$$C_{BS}(S_0, K, T, \sigma_{imp}) = C_{Heston}(K,T).$$

Thus, the *Heston-implied surface* is the mapping:

$$(K,T) \mapsto \sigma_{imp}(K,T).$$

Since the Heston call price is obtained from:

$$C = S_0 e^{-qT} P_1 - K e^{-rT} P_2,$$

with P_1, P_2 computed via Fourier inversion, the implied volatility surface is fully determined by the five model parameters:

$$(\kappa, \theta, \sigma, \rho, v_0).$$

2.2.7 Structure of the Volatility Surface Under Heston

Each parameter affects the geometry of the implied volatility surface in a precise way:

1. Correlation ρ controls the *skewness*

Negative correlation ($\rho < 0$) produces the familiar downward-sloping skew, typical in equity markets.

Mathematically, ρ enters the characteristic function through the terms $-\rho\sigma i\phi$ and $(\rho\sigma i\phi - b_j)^2$, modifying the asymmetry of the return distribution.

2. Volatility of volatility σ controls the *curvature (smile)*

Large σ increases **kurtosis**, leading to fatter tails and a more pronounced smile.

It appears quadratically in:

$$d = \sqrt{(\rho\sigma i\phi - b_j)^2 - \sigma^2(2u_j i\phi - \phi^2)}$$

3. Mean reversion speed κ **and long-run variance** θ

They control how volatility evolves in the term structure:

- High $\kappa \to$ faster reversion $\to$ short-dated options dominated by v_0, long-dated by θ
- Low $\kappa \to$ more persistent volatility $\to$ stronger term structure

This follows from the variance dynamics:

$$dv_t = \kappa(\theta - v_t)\,dt + \sigma\sqrt{v_t}\,dW_{2t}.$$

4. Initial variance v_0 controls short-term ATM vol

For $T \to 0$, the implied volatility converges to:

$$\sigma_{\text{imp}}(K = S_0, T) \longrightarrow \sqrt{v_0}.$$

2.2.8 Calibrating the Heston Model to the Market Surface

To match market implied volatilities, one must solve the optimization problem:

$$\min_{\kappa,\theta,\sigma,\rho,v_0} \sum_{i,j} w_{ij} \left(\sigma_{\text{imp}}^{Heston}(K_i, T_j) - \sigma_{\text{imp}}^{market}(K_i, T_j)\right)^2.$$

Where:

- (K_i, T_j) are the grid points of the implied volatility surface
- w_{ij} are optional weights (e.g., vega weighting)

The implied volatilities σ_{imp}^{Heston} are obtained by:

1. computing the Heston price
2. computing the Black–Scholes implied volatility

The Heston model is thus **parametric**: the entire volatility surface emerges from only **five parameters**.

2.3 Practical application

2.3.1 Mathematical Steps to Generate a Heston Volatility Surface

For each strike K and maturity T:

1. Compute characteristic function

$$f_j(x, v, T; \phi) = \exp\big(C_j(T;\phi) + D_j(T;\phi)v + i\phi x\big)$$

with:

$$x = \ln(S_0).$$

2. Compute probabilities P_j

$$P_j = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \Re\left[\frac{e^{-i\phi \ln K} f_j(\phi)}{i\phi}\right] d\phi.$$

3. Compute option price

$$C = S_0 e^{-qT} P_1 - K e^{-rT} P_2.$$

4. Invert Black–Scholes

   Solve for σ_{imp} such that:

$$C_{BS}(S_0, K, T, \sigma_{\text{imp}}) = C.$$

5. Repeat on a grid

$$\sigma_{\text{imp}}(K, T) \quad \forall (K, T) \longrightarrow \text{Heston volatility surface.}$$

The following volatility surface is generated from S&P 500 index options observed on 12 June 2023, using the calibrated parameters and plotted as a function of maturity and log-moneyness.

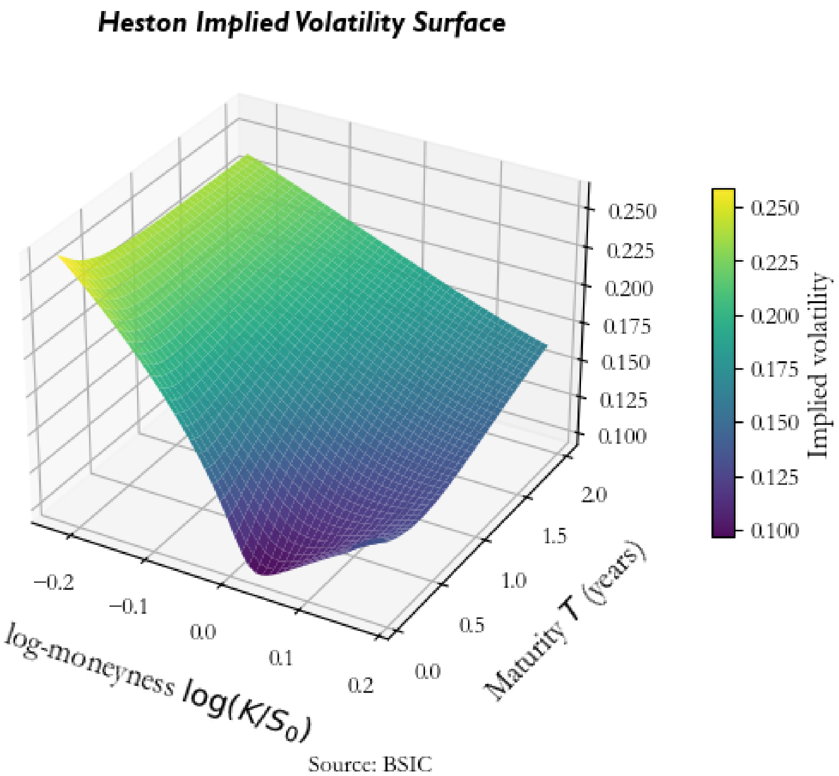

Figure 1: Heston implied volatility surface generated from S&P 500 index options.

3 SVI

3.1 Introduction

In option pricing, a **parametric model for implied volatility** offers an efficient and flexible way to describe how volatility behaves across different strikes and maturities.

Rather than assuming a constant volatility, as in the Black–Scholes framework, parametric models express the **implied volatility as an analytical function** of strike and time to maturity. If this function is smooth, its derivatives can also be computed analytically, which greatly reduces computational cost and simplifies calibration. Such models allow traders and quants to **interpolate and extrapolate** implied volatilities, enabling the pricing of options that are not directly quoted on the market.

We focus on **SVI parameterization**, as it provides a **simple yet powerful representation of the implied volatility surface**. Developed by Jim Gatheral at Merrill Lynch (1999), SVI captures the key empirical features of market smiles while allowing **analytical control over arbitrage conditions**. To continue the explanation of the model, three equivalent versions of the SVI parameterization will be introduced: the raw, the natural and the jump–wing.

3.2 Theory behind

3.2.1 The Raw SVI Parametrization

The **Stochastic Volatility Inspired (SVI)** model provides a simple analytical way to describe the **implied volatility smile** for a fixed maturity.

Instead of modeling volatility directly, SVI defines the **total implied variance** $w_{\text{imp}}^{\text{SVI}}(x)$ as a function of **log-moneyness** $x = \ln(K/F)$:

$$w_{\text{imp}}^{\text{SVI}}(x) = a + b\left[\rho(x-m) + \sqrt{(x-m)^2 + \sigma^2}\right],$$

where the parameter set is $\{a, b, \rho, m, \sigma\}$.

Parameter	**Description**
a	Overall vertical shift — determines the minimum variance level
b	Slope factor — controls the overall steepness of the smile
ρ	Skewness parameter — defines the asymmetry between call and put wings
m	Horizontal shift — moves the smile center left or right
σ	Curvature — controls the width and smoothness of the smile

Note: The parameter σ in SVI **does not represent** the volatility of the underlying asset, it is purely a shape parameter for the implied variance curve.

3.2.2 The Natural SVI Parametrization

$$w_{\text{imp}}^{\text{SVI}}(x) = \Delta + \frac{\omega}{2}\left[1 + \eta\rho(x-\mu) + \sqrt{(\eta(x-\mu)+\rho)^2 + (1-\rho^2)}\right].$$

This is the functional form derived as the limit of the Heston model for long maturities.

Parameters have the following interpretations:

Parameter	Description
Δ	Vertical shift of the curve
μ	Horizontal shift
ρ	Correlation between the underlying price and variance process
ω, η	Shape and scaling parameters linked to Heston dynamics

Although theoretically elegant, this form is **less practical for calibration**, and thus less used in implementation.

3.2.3 Parameter Bounds and Practical Constraints

When calibrating the SVI model, it is important to restrict the parameters to realistic and computationally stable ranges.

These **bounds** ensure that the optimization process converges to meaningful and arbitrage-free solutions.

Empirical and Numerical Constraints To maintain a **positive curvature** of the volatility smile (which is observed in real markets):

$$\sigma > 0.$$

To ensure the **slope of the smile** remains reasonable:

$$b \geq 0.$$

To avoid unrealistically high levels of implied variance:

$$a \leq \max_i \{w_i\},$$

where w_i are the observed total implied variances from market data.

The **center of the smile** should remain within the range of observed moneyness values:

$$2 \min_i \{x_i\} \leq m \leq 2 \max_i \{x_i\}.$$

Finally, since ρ represents the correlation between the Brownian motions driving the asset and its variance process, it must stay within:

$$\rho \in [-1, 1].$$

These conditions help maintain both **numerical stability** and **economic interpretability** of the fitted parameters.

3.2.4 Connection to the Natural Parametrization

The **Raw** and **Natural** forms of the SVI are mathematically equivalent and can be converted through the following relations:

$$(a, b, \rho, m, \sigma) = \left(\Delta + \frac{\omega}{2}(1-\rho^2),\ \frac{\omega\eta}{2},\ \rho,\ \mu - \frac{\rho}{\eta},\ \frac{\sqrt{1-\rho^2}}{\eta}\right),$$

$$(\Delta, \mu, \rho, \omega, \eta) = \left(a - \frac{\omega}{2}(1-\rho^2),\ m + \frac{\rho\sigma}{\sqrt{1-\rho^2}},\ \rho,\ \frac{2b\sigma}{\sqrt{1-\rho^2}},\ \frac{\sqrt{1-\rho^2}}{\sigma}\right).$$

These transformations allow the **Raw parameters** to be interpreted in terms of the **Heston model's stochastic volatility parameters**, linking the practical implementation of SVI to its theoretical foundation.

3.2.5 Jump–Wing (JW) Parametrization

The **Jump–Wing (JW)** parametrization rewrites the SVI model in terms of quantities that are **more intuitive for traders** and directly linked to observable features of the volatility smile.

It is defined in terms of the **Raw parameters** (a, b, ρ, m, σ) as follows:

$$v_\tau = \frac{a + b(-\rho m + \sqrt{m^2+\sigma^2})}{\tau},$$

$$\psi_\tau = \frac{b}{2\sqrt{w_\tau}}\left(\rho - \frac{m}{\sqrt{m^2+\sigma^2}}\right),$$

$$p_\tau = \frac{b(1-\rho)}{\sqrt{w_\tau}},$$

$$c_\tau = \frac{b(1+\rho)}{\sqrt{w_\tau}},$$

$$\hat{v}_\tau = \frac{1}{\tau}\left(a + b\sigma\sqrt{1-\rho^2}\right),$$

where $w_\tau = v_\tau \tau$ is the **total at-the-money implied variance**.

Parameter	Description
v_τ	At-the-money implied variance — the variance level of the smile at $(x = 0)$
ψ_τ	At-the-money implied volatility skew — measures the slope of the smile at the money
p_τ	Put wing slope — controls the steepness of the left side of the smile (out-of-the-money puts)
c_τ	Call wing slope — controls the steepness of the right side of the smile (out-of-the-money calls)
$\hat{v}_\tau$	Minimum variance level — the lowest point of the total variance curve

3.3 Practical application

3.3.1 SVI Parameter Calibration

The calibration of the **SVI model** aims to find the set of parameters (a, b, ρ, m, σ) that best fit market-observed implied volatilities. It will be described a systematic method for estimating these parameters from data.

The approach presented follows the ideas of Gatheral and Zeliade Systems (2009) and builds upon the parameter constraints discussed earlier.

Calibration Procedure Overview The calibration process involves three main steps:

1. **Define parameter bounds** — ensure realistic and numerically stable parameter ranges.
2. **Introduce an efficient reparametrization** — the **Quasi-Explicit (QE)** form, which simplifies optimization.
3. **Optimize the parameters** — using a numerical method such as the **Nelder–Mead algorithm**.

The objective is to minimize the error between the model's total implied variance $w_{\text{imp}}^{\text{SVI}}(x)$ and the market-observed total implied variances w_i^{market}:

$$\min_{a,b,\rho,m,\sigma} \sum_i \left[w_{\text{imp}}^{\text{SVI}}(x_i) - w_i^{\text{market}} \right]^2 .$$

3.3.2 Important Observations

From the raw SVI formula:

$$w_{\text{imp}}^{\text{SVI}}(x) = a + b \left[\rho(x - m) + \sqrt{(x - m)^2 + \sigma^2} \right],$$

we note that:

- When $\rho^2 \neq 1$, the function has a **unique minimum** at

$$w_{\min} = a + b\sigma\sqrt{1 - \rho^2}.$$

- When $\sigma \to 0$, the function becomes **piecewise linear**, making calibration unstable.
- To avoid such ill-posed cases, **additional parameter bounds** are imposed (e.g., ensuring $\sigma > 0$, $b > 0$).

These precautions ensure the optimization algorithm converges to a meaningful and unique solution.

3.3.3 Lower Bounds

1. The parameter a must be non-negative: $a \geq 0$.
2. The curvature parameter σ must be strictly positive: $\sigma \geq \sigma_{\min} > 0$.

Here, $\sigma_{\min}$ is a small positive constant (user-defined, typically very small).

3.3.4 Upper Bounds

There is no theoretical upper limit for σ, but in practice it is limited to keep the smile within realistic variance levels:

$$\sigma \leq 10.$$

(This value can be adjusted depending on computational limits.)

3.3.5 Bound on b (Smile Steepness)

From the no-dynamic-arbitrage condition

$$\left|\partial_x w_{\text{imp}}(x)\right| \leq 4, \quad \forall x, \forall \tau,$$

we obtain the practical upper bound:

$$b \leq \frac{4}{\tau(1+|\rho|)}.$$

Summary of All Bounds

$$\begin{gathered} 0 \leq a \leq \max_i \{w_i\}, \\ 0 \leq b \leq \frac{4}{\tau(1+|\rho|)}, \\ -1 \leq \rho \leq 1, \\ 2 \min_i \{x_i\} \leq m \leq 2 \max_i \{x_i\}, \\ \sigma_{\min} \leq \sigma \leq 10. \end{gathered}$$

These constraints ensure that:

- the volatility smile has a realistic shape,
- the optimization algorithm converges efficiently,
- and the fitted SVI parameters remain within economically meaningful ranges.

3.3.6 A New Parameterization — The Quasi-Explicit (QE) Form

The SVI model has **five parameters**, which makes calibration computationally expensive.

If we can reduce the dimensionality of the problem, the calibration process becomes faster and more efficient.

A convenient reparameterization, called the **Quasi-Explicit (QE) parameterization**, achieves exactly that.

Step 1: Change of Variable We define a new variable:

$$y(x) = \frac{x - m}{\sigma}.$$

This substitution simplifies the expression of the total implied variance in the **raw SVI form**.

Step 2: Rewriting the Variance Equation Starting from the raw parameterization:

$$w_{\text{imp}}^{\text{SVI}}(x) = a + b\left[\rho(x-m) + \sqrt{(x-m)^2 + \sigma^2}\right],$$

and substituting $y(x)$, we can rewrite it as:

$$w_{\text{imp}}^{\text{SVI}}(x) = a + b\sigma\left[\rho y(x) + \sqrt{y(x)^2 + 1}\right].$$

This can be expressed more compactly as a **linear combination**:

$$w_{\text{imp}}^{\text{SVI}}(x) = \hat{a} + d\,y(x) + c\,z(x),$$

where

$$\hat{a} = a, \quad d = b\rho\sigma, \quad c = b\sigma, \quad z(x) = \sqrt{y(x)^2 + 1}.$$

Step 3: Inner and Outer Parameters This reparameterization divides the parameters into two groups:

- **Inner parameters** $\to \hat{a}, d, c$ — enter linearly in the equation.
- **Outer parameters** $\to \sigma, m$ — enter nonlinearly.

This "split" greatly simplifies the optimization.

Why It Matters

- Reduces a **5-dimensional** nonlinear optimization to a **2-dimensional** one.
- Makes calibration **quicker and less prone to local minima**.
- The method is directly inspired by Zeliade Systems (2009), where this QE form was first proposed.

3.3.7 Optimization

After defining the parameter bounds, the next step is to calibrate the SVI model by minimizing the difference between the model-generated and market-observed total implied variances.

In the **Quasi-Explicit (QE) parameterization**, the inner parameters $(\hat{a}, d, c)$ are optimized within a **compact and convex domain** defined by the following bounds:

$$\mathcal{D} = \left\{ (\hat{a}, d, c) : \; 0 \le c \le 4\sigma, \; |d| \le c, \; |d| \le 4\sigma - c, \; 0 \le \hat{a} \le \max_i \{w_i\} \right\}.$$

Here, w_i are the market-observed total implied variances.

For any fixed pair of outer parameters (σ, m), the optimization problem becomes:

$$\min_{(\hat{a},d,c)\in\mathcal{D}} f_{x_i,w_i}(\hat{a}, d, c),$$

where the **quadratic cost function** is defined as:

$$f_{x_i,w_i}(\hat{a}, d, c) = \sum_{i=1}^{n} \left[\hat{a} + d\, y(x_i) + c\, z(x_i) - w_i\right]^2.$$

Key Properties

- The function f_{x_i,w_i} is **smooth and convex**, which guarantees a **unique minimum** for every fixed (σ, m) within the convex set $\mathcal{D}$.
- The **gradient of** f is linear in $(\hat{a}, d, c)$, so the solution can be found **analytically** by inverting a simple 3×3 **linear system**. This makes the inner optimization **extremely fast**.

Two-Level Optimization Structure The calibration process is divided into two nested loops:

1. **Inner optimization** $\rightarrow$ solves for the best $(\hat{a}, d, c)$ for fixed (σ, m).
2. **Outer optimization** $\rightarrow$ adjusts (σ, m) to minimize the total fitting error.

At each iteration of the outer loop, the inner loop computes the analytical minimum of f_{x_i,w_i}, returning the corresponding **raw SVI parameters** (a^*, b^*, ρ^*).

Result: This structure allows the algorithm to **drastically reduce computation time** while maintaining accuracy in fitting the SVI surface to real market data.

As a final illustration of the SVI framework, we report a practical calibration to S&P 500 index option data with time to maturity $T \approx 0.51$ years. In Figure 2, the black dots represent the market implied volatilities, while the solid line shows the corresponding SVI fit, both plotted as a function of log–moneyness $m = \log(K/S_0)$.

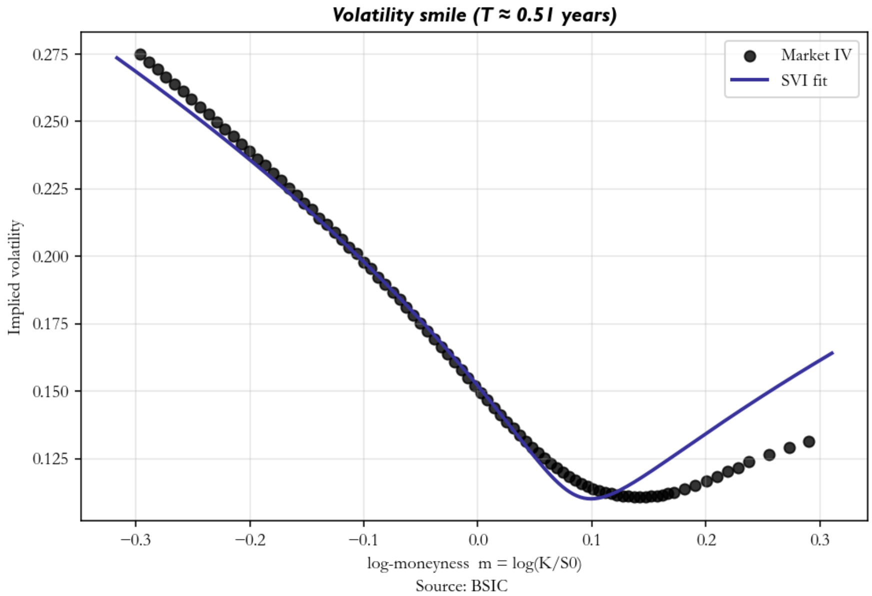

Figure 2: SVI volatility smile generated from S&P 500 index options.

4 Local Volatility

4.1 Introduction: Local Volatility

The local volatility model advances option pricing by replacing the Black–Scholes model's constant, global volatility assumption with a deterministic function of spot and time, $\sigma_{\text{loc}}(S, t)$. This refinement allows the instantaneous volatility to depend on where the underlying actually is and when, addressing persistent features in real markets such as skews and smiles, which a constant-volatility model cannot capture.

What is "local" about local volatility?

Whereas Black–Scholes implied volatility, $\sigma_{IV}(K, T)$, is obtained for each individual option (characterized by strike K and expiry T), it does not reflect a true instantaneous or path-specific market view. Instead, it acts as a weighted average, condensing all the possible, potentially time- and state-dependent volatilities encountered by the underlying as it evolves along every path that would lead to the option being in or out of the money at expiry. Implied volatility reflects the market's consensus for "the average total risk" embedded in the value of an option given its payoff structure, not the actual risk at any particular spot or time.

Local volatility, by contrast, answers the question:

> If the underlying is precisely at level S at time t, what is the market's consensus about its instantaneous volatility at that exact point?

The local volatility function, $\sigma_{\text{loc}}(S, t)$, thus provides a map of pointwise, time-conditional volatility, in effect, the "velocity" the market expects if and when the underlying finds itself at a given state and time, regardless of how it got there.

How is local volatility constructed?

The key to recovering local volatility is the entire set of market vanilla option prices, typically quoted as implied volatilities $\sigma_{IV}(K, T)$ across a grid of strikes and maturities. Provided these prices are sufficiently smooth and arbitrage-free, results by Dupire and subsequent researchers show that you can reconstruct a unique local volatility surface via the so-called Dupire formula, translating the averaged market information into a deterministic and non-parametric model mechanism. Mathematically, this involves differentiating prices (or implied vols) with respect to strike K and maturity T, then evaluating the resulting local volatility at the point $(S = K, t = T)$.

Why does local volatility matter?

- It provides the flexibility to fit all observed vanilla option prices at once, achieving full static consistency with the market.
- It enables coherent pricing and hedging of exotic or path-dependent options, since every state and time has a prescribed volatility.
- It offers a practical, Markovian (memoryless) model suitable for simulation and risk analysis, directly linking observed smiles and skews in the market to model dynamics.

4.2 Theory behind

4.2.1 The Local Volatility Model Framework

We model the underlying asset price S_t under the risk-neutral measure with dynamics

$$dS_t = (r - q)S_t\,dt + \sigma_{\text{loc}}(S_t, t)S_t\,dW_t$$

where:

- r is the risk-free rate.
- q the dividend or foreign yield.
- $\sigma_{\text{loc}}(S, t)$ is a deterministic, time- and state-dependent local volatility function.
- W_t is a standard Brownian motion.

The goal is to find σ_{loc} such that the model exactly reproduces the observed vanilla option prices for all strikes K and expiries T.

4.2.2 Theoretical Foundations

The option price $V(t, S)$ satisfies the backward partial differential equation (PDE):

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2_{\text{loc}}(S, t)S^2\frac{\partial^2 V}{\partial S^2} + (r - q)S\frac{\partial V}{\partial S} - rV = 0$$

with terminal condition $V(T, S) = \Phi(S)$ given by the payoff.

Considering the market call price surface $C(0; K, T)$ as a function of strike and maturity, Dupire's forward PDE reads:

$$\frac{\partial C}{\partial T} = \frac{1}{2}\sigma^2_{\text{loc}}(K, T)K^2\frac{\partial^2 C}{\partial K^2} - (r - q)K\frac{\partial C}{\partial K} - qC$$

Rearranged, this gives the Dupire local volatility formula

$$\sigma^2_{\text{loc}}(K, T) = \frac{2\left(\partial_T C + (r - q)K\partial_K C + qC\right)}{K^2\partial_{KK}C}$$

This expresses local volatility entirely in terms of observable vanilla option prices and their derivatives with respect to strike and maturity.

4.2.3 Practical Considerations

For numerical stability, it is convenient to work in normalized variables: the moneyness $x = \frac{K}{F(0,T)}$ and discounted call prices:

$$\widehat{C}(T, x) = \frac{C(0; K, T)}{P(0, T)\,F(0, T)}\,.$$

Here $F(0,T) = S_0 e^{(r-q)T}$ is the forward price, and $P(0,T)$ is the discount factor. Dupire's equation then simplifies to:

$$\partial_T \widehat{C} = \frac{1}{2}\widehat{\sigma}^2_{\text{loc}}(T,x)\, x^2 \partial_{xx}\widehat{C}, \quad \text{with} \quad \widehat{\sigma}^2_{\text{loc}}(T,x) = \frac{2\,\partial_T \widehat{C}}{x^2 \partial_{xx}\widehat{C}}.$$

This normalization eliminates explicit dependence on r and q and aligns the numerical grid with market forwards.

4.2.4 From Implied Volatility to Local Volatility

Market quotes imply Black–Scholes volatilities $\sigma_{IV}(K,T)$ for each strike and maturity. Using chain rule and Black–Scholes sensitivities, we obtain

$$\sigma^2_{\text{loc}}(K,T) = \frac{\sigma^2_{IV} + 2T\,\sigma_{IV}\,\frac{\partial \sigma_{IV}}{\partial T} + 2(r-q)KT\,\sigma_{IV}\,\frac{\partial \sigma_{IV}}{\partial K}}{1 + 2d_1 K\sqrt{T}\frac{\partial \sigma_{IV}}{\partial K} + K^2 T\left[d_1 d_2\left(\frac{\partial \sigma_{IV}}{\partial K}\right)^2 + \sigma_{IV}\frac{\partial^2 \sigma_{IV}}{\partial K^2}\right]}$$

where d_1, d_2 are computed with σ_{IV}. This formula links the implied volatility surface to the local volatility function.

4.2.5 Interpretation and Smile Dynamics

- The local volatility $\sigma_{\text{loc}}(S,t)$ represents the market-implied instantaneous volatility at price S and time t.
- Because σ_{loc} is deterministic, the model predicts "sticky-strike" dynamics: the implied volatility smile remains fixed relative to strikes when spot moves.
- This contrasts with empirical market behavior, which often follows "sticky-delta" dynamics, indicating that local volatility is a baseline static model rather than a full stochastic volatility solution.
- Local volatility surfaces can be sensitive to data quality and interpolation methods, especially at extreme strikes (wings). Ensuring smoothness and arbitrage-freeness remains critical.

4.3 Practical application

To conclude, we show an example of the local volatility surface implied by the calibrated option data. Starting from a smooth implied–volatility surface for S&P 500 index options observed on 12 June 2023, the local volatility $\sigma_{\text{loc}}(K, T)$ is obtained via the Dupire formula and plotted as a function of maturity and log–moneyness.

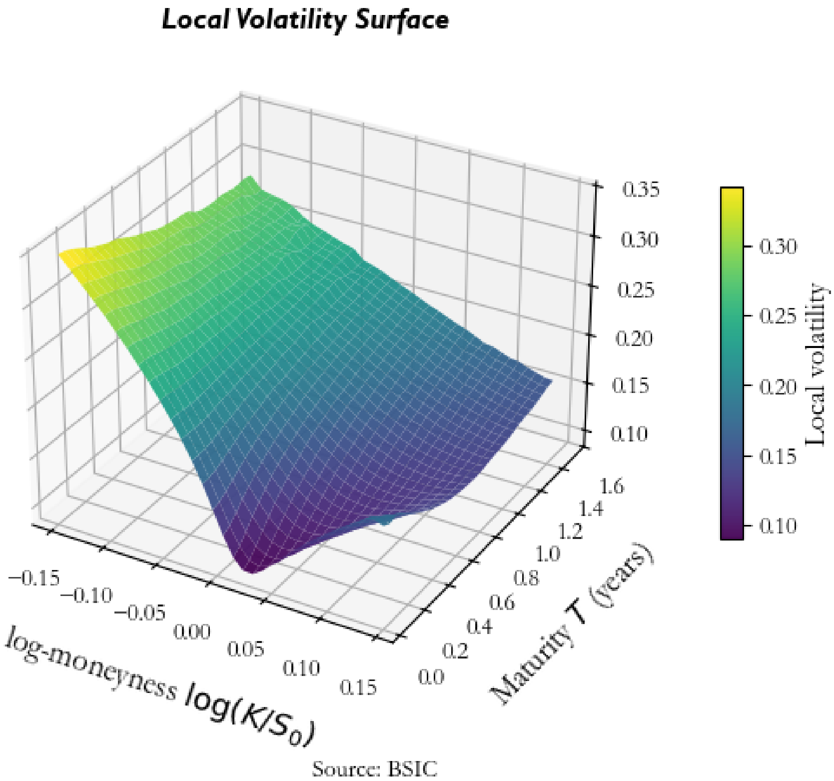

Figure 3: Local volatility surface generated from S&P 500 index options.

5 Vanna–Volga

5.1 Introduction: Vanna–Volga

The Vanna–Volga (VV) method is a widely adopted, non-parametric approach in foreign exchange (FX) markets for constructing smile-consistent implied volatility surfaces and pricing "first-generation" exotic options. Instead of calibrating complex stochastic models, VV starts directly from a handful of liquid market-implied volatilities for each expiry, specifically, the at-the-money (ATM) volatility σ_{ATM}, the 25-delta risk reversal (RR), and the 25-delta butterfly (BF). These three instruments summarize the overall level, skew (asymmetry), and smile (curvature) of the market surface.

The method's underlying principle is to replicate any arbitrary-strike option by combining three benchmark options (typically ATM, 25-delta call, and 25-delta put), whose volatilities are directly observed in the market. This portfolio is locally vega-neutral and designed so that its price sensitivity matches the market's volatility smile through "vanna" (sensitivity to spot and vol) and "volga" (sensitivity to vol squared). The cost of constructing and hedging this portfolio under market and Black–Scholes volatilities is compared, and the resultant adjustment is added to the vanilla Black–Scholes price to yield a more market-consistent value.

Specifically, the VV method does not require a global volatility parameter, nor does it perform costly multi-dimensional calibrations. Instead, it interpolates the smile using these three points and adjusts the Black–Scholes model accordingly. This means that for any given strike K and expiry T, the final implied volatility $\sigma_{VV}(K, T)$ will exactly match the market's ATM, RR, and BF inputs at their respective strikes and interpolate smoothly across other strikes.

While the method is intuitive and computationally light (making it ideal for booking and risk-managing exotics in live trading environments), it is not without limitations. The standard VV approach may not extrapolate well in the extreme wings of the smile or under unusually strong skew conditions, sometimes requiring theoretical corrections or smoothing.

Originating in FX, the VV framework has seen refinements by practitioners such as Lipton, McGhee, Wystup, and Castagna–Mercurio, but it remains relatively rare in equity/commodity markets.

5.2 Theory behind

The VV method is essentially a way to adjust the Black–Scholes option price so that it reflects the market's volatility smile. It uses three liquid market quotes: the ATM volatility, the 25-delta risk reversal, and the 25-delta butterfly as the foundation. These three points capture the level, skewness and curvature of the implied volatility smile.

In particular, it aligns exposures to volatility (vega), the change of vega with spot (vanna), and the change of vega with volatility (volga). Once this portfolio is set up, the cost of hedging with market quotes is compared to the theoretical Black–Scholes cost. The difference becomes the adjustment added to the Black–Scholes price, yielding a smile-consistent option value.

In practice, this means the VV method takes the easy-to-compute Black–Scholes price and "corrects" it using real market information. By VV we obtain a set of option prices and implied volatilities that match the three input quotes exactly and interpolate smoothly across strikes. This approach is simple, intuitive and model-free.

Because of its simplicity, VV is not perfect: it can behave poorly in the extreme wings of the smile (very deep ITM or OTM strikes) or under extreme skew.

5.2.1 Vega, Vanna and Volga

Before diving into the method, we introduce Vega, Vanna and Volga.

Vega. Vega ($\mathcal{V}$) quantifies the sensitivity of an option's price to changes in implied volatility. For a European option with current spot S_0, volatility σ, time to maturity $T-t$, and continuously compounded yield q:

$$\text{Vega} = S_0\sqrt{T-t}N'(d_1)e^{-q(T-t)}$$

where

$$N'(d_1) = \frac{1}{\sqrt{2\pi}}e^{-d_1^2/2}$$

and

$$d_1 = \frac{\ln(S_0/K) + (r - q + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}.$$

Vanna. Vanna represents the risk to the skew increasing. It is used to monitor the Vega exposure. It can be defined in three different ways:

$$\frac{\partial V}{\partial \sigma}, \quad \frac{\partial^2 P}{\partial \sigma^2}.$$

Vanna is derived by

$$\text{Vanna} = e^{-q(T-t)}\sqrt{T-t}N'(d_1)\left(\frac{d_2}{\sigma}\right)$$

In terms of Vega, it can be written as

$$\text{Vanna} = \text{Vega} \cdot \frac{d_2}{S\sigma}$$

where $d_2 = d_1 - \sigma\sqrt{T-t}$.

Volga. Volga represents the sensitivity of Vega with respect to the change in volatility and shows the risk to the smile becoming more pronounced. It measures the convexity of option price with respect to volatility. Vega–Volga is the same relationship we have with Gamma–Delta. It can be defined in two different ways:

$$\frac{\partial V}{\partial \sigma}, \quad \frac{\partial^2 P}{\partial \sigma^2}.$$

Volga is derived as

$$\text{Volga} = e^{-q(T-t)}\sqrt{T-t}N'(d_1)\left(\frac{d_1 d_2}{\sigma}\right)$$

In terms of Vega, it can be written as

$$\text{Volga} = \text{Vega} \cdot \frac{d_1 d_2}{S\sigma}.$$

5.2.2 The Vanna–Volga Option Pricing Formula

The Vanna–Volga option price $C^{VV}(K)$ is obtained by adding to the Black–Scholes theoretical price $C^{BS}(K)$ the cost difference of the hedging portfolio induced by the market implied volatilities with respect to the constant volatility σ:

$$C^{VV}(K) = C^{BS}(K) + \sum_{i=1}^{3} x_i(K)\big(C^M(K_i) - C^{BS}(K_i)\big)$$

where $C^M(K)$ denotes the observed market call option price for strike K.

The first step in VV is to build a portfolio of three options with same maturity but different strikes, so that the portfolio can hedge the price variation of the call $C(K)$ up to the second order in the underlying and the volatility. Under diffusion dynamics both S_t and σ_t, by Itô's lemma we have

$$\begin{aligned}
&dC^{BS}(K) - \Delta\, dS_t - \Delta\, \delta S_t\, dt - \sum_{i=1}^{3} x_i\, dC^{BS}(K_i) \\
&= \left[\frac{\partial C^{BS}(K)}{\partial t} - \Delta\, \delta S_t - \sum_{i=1}^{3} x_i \frac{\partial C^{BS}(K_i)}{\partial t}\right] dt \\
&+ \left[\frac{\partial C^{BS}(K)}{\partial S} - \Delta - \sum_{i=1}^{3} x_i \frac{\partial C^{BS}(K_i)}{\partial S}\right] dS_t \\
&+ \left[\frac{\partial C^{BS}(K)}{\partial \sigma} - \sum_{i=1}^{3} x_i \frac{\partial C^{BS}(K_i)}{\partial \sigma}\right] d\sigma_t \\
&+ \frac{1}{2}\left[\frac{\partial^2 C^{BS}(K)}{\partial S^2} - \sum_{i=1}^{3} x_i \frac{\partial^2 C^{BS}(K_i)}{\partial S^2}\right] (dS_t)^2 \\
&+ \frac{1}{2}\left[\frac{\partial^2 C^{BS}(K)}{\partial \sigma^2} - \sum_{i=1}^{3} x_i \frac{\partial^2 C^{BS}(K_i)}{\partial \sigma^2}\right] (d\sigma_t)^2 \\
&+ \left[\frac{\partial^2 C^{BS}(K)}{\partial S \partial \sigma} - \sum_{i=1}^{3} x_i \frac{\partial^2 C^{BS}(K_i)}{\partial S \partial \sigma}\right] dS_t\, d\sigma_t.
\end{aligned}$$

We zero out coefficients dS_t, $d\sigma_t$, $(d\sigma_t)^2$ and $dS_t d\sigma_t$ so that no stochastic terms are involved in its differential. Applying the Black–Scholes PDE, we get

$$dC^{BS}(K) - \Delta_t dS_t - \Delta_t \delta S_t dt - \sum_{i=1}^{3} x_i dC_i^{BS}(K_i) = r\left(C^{BS}(K) - \Delta_t S_t - \sum_{i=1}^{3} x_i C_i^{BS}(K_i)\right) dt.$$

Although volatility is stochastic and options are priced by the Black–Scholes formula, we can still have a perfect hedge.

Assuming Delta-hedging and that the replicating portfolio is Vega-neutral and Gamma-neutral, we can find the weights by imposing same Vega, Vanna and Volga:

$$\begin{aligned}
\frac{\partial C^{BS}}{\partial \sigma}(K) &= \sum_{i=1}^{3} x_i(K)\,\frac{\partial C^{BS}}{\partial \sigma}(K_i),\\
\frac{\partial^2 C^{BS}}{\partial \sigma^2}(K) &= \sum_{i=1}^{3} x_i(K)\,\frac{\partial^2 C^{BS}}{\partial \sigma^2}(K_i),\\
\frac{\partial^2 C^{BS}}{\partial \sigma\,\partial S_0}(K) &= \sum_{i=1}^{3} x_i(K)\,\frac{\partial^2 C^{BS}}{\partial \sigma\,\partial S_0}(K_i).
\end{aligned}$$

By solving equations we derive the unique solution of the weights

$$x_1(K) = \frac{\mathcal{V}(K)}{\mathcal{V}(K_1)}\frac{\ln\frac{K_2}{K}\,\ln\frac{K_3}{K}}{\ln\frac{K_2}{K_1}\,\ln\frac{K_3}{K_1}},$$

$$x_2(K) = \frac{\mathcal{V}(K)}{\mathcal{V}(K_2)}\frac{\ln\frac{K}{K_1}\,\ln\frac{K_3}{K}}{\ln\frac{K_2}{K_1}\,\ln\frac{K_3}{K_2}},$$

$$x_3(K) = \frac{\mathcal{V}(K)}{\mathcal{V}(K_3)}\frac{\ln\frac{K}{K_1}\,\ln\frac{K}{K_2}}{\ln\frac{K_3}{K_1}\,\ln\frac{K_3}{K_2}}.$$

If $K = K_j$, then $x_i(K) = 1$ for $i = j$ and zero otherwise.

The Vanna–Volga price preserves convexity and matches probability mass by construction. The second derivative with respect to strike, $\frac{\partial^2 C}{\partial K^2}$, remains non-negative, reflecting positive risk-neutral density.

5.2.3 The 1st and the 2nd Approximation of Vanna–Volga Implied Volatility

We want an implied volatility $\varrho(K)$ for any strike K that is consistent with the VV price. Instead of solving for $\varrho(K)$ by numerically inverting Black–Scholes each time, we can find a closed-form approximation by Taylor-expanding the VV price around a reference vol, in this case $\sigma = \sigma_2$ which is the ATM vol.

The first-order approximation is given by:

$$\varrho(K) = \mathcal{X}_1(K)\sigma_1 + \mathcal{X}_2(K)\sigma_2 + \mathcal{X}_3(K)\sigma_3$$

Implied vol at strike K is a weighted average of the three market quotes $\sigma_1, \sigma_2, \sigma_3$. We have that $\mathcal{X}_i(K)$ depend on log-strike ratios and they sum to 1. It produces a quadratic function of log-strike. It is very simple, fast and intuitive, but overestimates/underestimates vol at far ITM/OTM strikes.

The second-order approximation is given by

$$\varrho(K) = \sigma_2 \; + \; \frac{-\sigma_2 + \sqrt{\sigma_2^2 + d_1(K)d_2(K)\left[2\sigma_2 D_1(K) + D_2(K)\right]}}{d_1(K)d_2(K)}.$$

where

$$\begin{aligned} D_1(K) = {} & \frac{\ln(K_2/K)\,\ln(K_3/K)}{\ln(K_2/K_1)\,\ln(K_3/K_1)}\,\sigma_1 + \frac{\ln(K/K_1)\,\ln(K_3/K)}{\ln(K_2/K_1)\,\ln(K_3/K_2)}\,\sigma_2 \\ & + \frac{\ln(K/K_1)\,\ln(K/K_2)}{\ln(K_3/K_1)\,\ln(K_3/K_2)}\,(\sigma_3 - \sigma_2), \end{aligned}$$

$$\begin{aligned} D_2(K) = {} & \frac{\ln(K_2/K)\,\ln(K_3/K)}{\ln(K_2/K_1)\,\ln(K_3/K_1)}\,d_1(K_1)d_2(K_1)(\sigma_1 - \sigma_2)^2 \\ & + \frac{\ln(K/K_1)\,\ln(K/K_2)}{\ln(K_3/K_1)\,\ln(K_3/K_2)}\,d_1(K_3)d_2(K_3)(\sigma_3 - \sigma_2)^2. \end{aligned}$$

The second approximation is not only accurate within the interval $[K_1, K_3]$ but also in the wings, even for extreme values of put Deltas.

6 SABR

6.1 Introduction

The SABR model is a parsimonious stochastic-volatility framework used pervasively in rates for caplets/floorlets and swaptions. It provides a compact, trader-friendly parameterization of the smile across strikes and tenors while retaining closed-form (asymptotic) implied-vol formulas for fast calibration and risk.

It is used in cap/floor markets, where smiles are mainly quoted in normal (Bachelier) or lognormal (Black) vols depending on regime, and in swaption markets, where smiles are quoted per expiry $\times$ underlying swap tenor, forming the vol cube.

SABR is used in rates since it captures level dependence of ATM vol (via β) and controls skew and curvature (via ρ, α).

6.2 Theory behind

6.2.1 Model definition

The SABR model is a two-factor model for the forward prices of the underlying, whose second factor (the spot volatility σ_t) is a stochastic process. Mathematically it can be defined as follows:

$$\begin{cases} dF_t = \sigma_t F_t^{\beta}\, dW_t \\ d\sigma_t = \alpha\sigma_t\, dB_t \end{cases} \quad \text{and} \quad \rho = \frac{d\langle W, B\rangle_t}{dt} \quad \text{for } t \in [0, T]$$

where F_t is the forward price of the underlying, W_t and B_t are Brownian motions and ρ is the correlation between them. α is a parameter which can be understood as the volatility of volatility ($\alpha > 0$), while β is a parameter $\beta \in [0, 1]$ that determines how at-the-money volatility changes when forward price changes.

β controls the relationship between volatility and price level, how "elastic" volatility is relative to the forward:

- If $\beta = 1$ then $dF_t = \sigma_t F_t dW_t$. This is the lognormal model (similar to Black–Scholes), where volatility is proportional to price level.
- If $\beta = 0$ then $dF_t = \sigma_t dW_t$. This is the normal model (Bachelier model), where volatility is independent of price level. Forward price can become negative (relevant for low or near-zero rate markets).
- If $\beta = \frac{1}{2}$ then we get a Cox–Ingersoll–Ross (CIR) like process, where volatility grows sublinearly with price level.

6.2.2 Hagan's Formula

Hagan's formula allows us to express the implied volatility σ_B of an option in terms of its strike price, time to maturity and the parameters previously mentioned from the SABR model:

$$\sigma_B(F_0,K) = \frac{\sigma_0}{(F_0K)^{\frac{1-\beta}{2}}\left[1+\frac{(1-\beta)^2}{24}\log^2\left(\frac{F_0}{K}\right)+\frac{(1-\beta)^4}{1920}\log^4\left(\frac{F_0}{K}\right)+\cdots\right]}\cdot\left(\frac{z}{x(z)}\right)$$
$$\cdot\left[1+\left(\frac{(1-\beta)^2}{24}\frac{\sigma_0^2}{F_0^{2(1-\beta)}}+\frac{1}{4}\frac{\rho\beta\alpha\sigma_0}{F_0^{1-\beta}}+\frac{2-3\rho^2}{24}\alpha^2\right)T\right]+\cdots$$

where σ_0 is the initial volatility and the z and $x(z)$ terms are given by the following expressions

$$z=\frac{\alpha}{\sigma_0}\left(\frac{F_0}{K}\right)^{\frac{1-\beta}{2}}\log\left(\frac{F_0}{K}\right), \qquad x(z)=\log\left(\frac{\sqrt{1-2\rho z+z^2}+z-\rho}{1-\rho}\right).$$

Moreover, the formula can be further simplified in the case of at-the-money options (where $K = F_0$) because of some logarithms vanishing:

$$\sigma_B(F_0,K)=\frac{\sigma_0}{F_0^{1-\beta}}\left(\frac{z}{x(z)}\right)\left[1+\left(\frac{(1-\beta)^2}{24}\frac{\sigma_0^2}{F_0^{2(1-\beta)}}+\frac{1}{4}\frac{\rho\beta\alpha\sigma_0}{F_0^{1-\beta}}+\frac{2-3\rho^2}{24}\alpha^2\right)T\right]+\cdots$$

As Hagan et al. indicate, the terms that account for "…" can be omitted because of their small size, even though this omission can lead to relative errors when pricing options. Alòs and Garcìa provide an approximation for both formulae when ignoring the small-sized terms and when $\beta = 1$:

$$\sigma_B(F_0,K)\approx\sigma_0\left[1+\left(\frac{\rho\alpha\sigma_0}{4}+\frac{2-3\rho^2}{24}\alpha^2\right)T\right]\left(\frac{z}{x(z)}\right),$$
$$\sigma_B(F_0,K)\approx\sigma_0\left[1+\left(\frac{\rho\alpha\sigma_0}{4}+\frac{2-3\rho^2}{24}\alpha^2\right)T\right].$$

The complete formulae and the approximations can show how the implied volatility depends on the main parameters of the model. Both α and ρ affect the implied volatility skew, as high values of α make the curvature increase, while high values of $|\rho|$ make the skew more pronounced.

One important result is that the short-term skew converges to a finite limit that depends on the product $\alpha\rho$. However, empirical market data often exhibits a blow-up in the short-end of the implied volatility surface, meaning that the skew tends to infinity as maturity approaches zero. Consequently, the SABR model cannot capture this short-end blow up.

Another notable issue with the SABR model is that Hagan's closed-form approximation is not arbitrage-free. The approximate density function implied by the model can take negative values for low strike prices (Alòs and Garcìa, 2021), potentially leading to arbitrage opportunities and inaccurate pricing results.

6.2.3 SABR Model Calibration

The goal of calibrating a model is to obtain parameters that allow reproducing market volatility surface for all strike prices and maturities.

However, as mentioned before, stochastic volatility models such as the SABR are not able to reproduce the dependence of the implied volatility with respect to time to maturity, so trying to obtain a set of parameters for all maturities would not be effective. A common practice in the financial industry is to calibrate the model obtaining a set of parameters for each fixed maturity, so we will use this approach for calibrating. Hence, we need to minimize the sum of squared errors to obtain a triple of (α, ρ, σ_0) for each time to maturity:

$$(\hat{\alpha}, \hat{\rho}, \hat{\sigma}_0)_T = \arg \min_{\alpha,\rho,\sigma_0} \sum_{K_T} \left(\sigma^{mkt}_{K,T} - \sigma^{SABR}(F_0, T, K, \alpha, \rho, \sigma_0)\right)^2$$

where $\sigma^{mkt}_{K,T}$ is the market implied volatility and σ^{SABR} is the SABR implied volatility for a given strike K and maturity T. In this context, the implied volatility of the SABR model is the implied volatility obtained through Hagan's formula, and we use Alòs and Garcìa approximation in the minimization problem as the gains of precision for including other terms are not significant.

7 Conclusions

This project has compared four complementary approaches to volatility surface modelling, Heston, SVI, local volatility, and Vanna–Volga, highlighting their structural differences, asset-class focus, and practical purposes. Heston (together with SABR in the rates space) belongs to the family of *stochastic-volatility models for the underlying dynamics*, defined via stochastic differential equations with a small number of parameters. These models are particularly suited to equity, FX, and interest-rate markets where one requires a dynamic description of paths and risk factors that is broadly consistent with observed smiles. In contrast, SVI, local volatility, and Vanna–Volga are primarily *surface- or price-level constructions*: SVI provides a parametric total-variance fit per expiry, local volatility recovers a deterministic $\sigma_{\text{loc}}(S,t)$ from the full grid of vanilla prices, and Vanna–Volga directly adjusts Black–Scholes prices from a handful of benchmark quotes.

From an asset-class perspective, Heston is widely used in equity and FX to generate stochastic-volatility surfaces and to interpret skew and smile in terms of mean reversion, volatility of volatility, and spot/vol correlation, while SABR plays an analogous role in interest-rate markets for cap/floor and swaption smiles. SVI has become a market standard for equity index and single-name volatility surfaces because it combines parsimony with explicit no-arbitrage constraints across log-moneyness and maturities, and its "natural" and "jump-wing" forms link directly to the long-maturity limit of Heston. Local volatility is most natural for equity and FX index books where static replication of exotics is important: by construction it fits the entire vanilla grid exactly, but its deterministic, sticky-strike dynamics make it a baseline rather than a fully realistic stochastic description. Finally, Vanna-Volga is predominantly an FX tool for first-generation exotics, where the market quotes smiles in terms of ATM, risk reversals, and butterflies, and desks need a fast, model-free way to embed this information into prices.

In terms of purpose, Heston and SABR are primarily used for *dynamic hedging, risk-factor simulation, and consistent pricing of both vanillas and path-dependent exotics*, accepting a coarser fit to the raw surface in exchange for a coherent stochastic structure. SVI's role is to provide a *front-office quality, arbitrage-free implied-volatility surface* that interpolates and extrapolates market quotes smoothly and robustly, serving as the backbone for quoting and risk systems. Local volatility is designed to achieve *full static replication of vanillas and Markovian pricing of exotics*, trading off realistic dynamics (sticky-delta behaviour, volatility clustering) for exact consistency with today's surface. Vanna–Volga is explicitly desk-oriented: it offers a quick, smile-consistent correction to Black–Scholes prices that matches a small set of liquid FX quotes exactly and preserves convexity, making it attractive for daily marking and risk of barrier, digital, and other simple exotics despite known limitations in the wings.

Overall, the analysis confirms that no single framework dominates across all use cases. Instead, the appropriate choice depends on the asset class, the required balance between fit and dynamics, and whether the desk prioritises theoretical consistency, calibration stability, or speed and interpretability in a live trading environment. Within this landscape, SVI and local volatility provide powerful tools for constructing and interrogating the volatility surface itself, while Heston and related stochastic-volatility models serve as workhorses for scenario generation, risk analysis, and pricing of more complex derivatives.

Model	Type	Typical asset class	Main purpose	Key limitations
Heston	Stochastic vol (SDE)	Equity, FX	Dynamic pricing/hedging; generating stochastic vol paths consistent with smiles.	Limited flexibility to fit full surface; calibration can be delicate.
SABR	Stochastic vol (SDE)	Rates	Modelling swaption and cap/floor smiles in a term-structure framework.	Asymptotic model; may struggle in extreme wings or stressed regimes.
SVI	Parametric vol surface	Equity, some FX	Smooth, arbitrage-aware fit of implied vol surface for quoting and risk.	Static only; no path dynamics; quality depends on calibration choices.
Local Volatility	Deterministic $\sigma_{loc}(S, t)$	Equity, FX indices	Exact fit of vanilla grid; Markovian pricing of exotics.	Sticky-strike dynamics; very sensitive to noise and arbitrage in input.
Vanna–Volga	Pricing adjustment	FX exotics	Fast smile-consistent correction to Black–Scholes for simple exotics.	Heuristic; no underlying dynamic model; weaker in wings/complex payoffs.

Table 1: Summary of the main models, their typical use, and key limitations.

References

Castagna, A. and Mercurio, F. (2007). The vanna-volga method for implied volatilities. Cutting Edge: Option Pricing.

Huang, K. (2012). Vanna Volga and smile-consistent implied volatility surface of equity index option. Working paper.

Wystup, U. (2017). Vanna-Volga pricing. MathFinance AG, Waldems, Germany. Available at `www.mathfinance.com`.

White, R. (2012). Local volatility. OpenGamma Quantitative Research.

Derman, E., Kani, I., and Zou, J. Z. (1996). The local volatility surface: Unlocking the information in index option prices. Goldman Sachs Quantitative Strategies Research Notes.

Magnusson, E. (2010). Implied volatility surface construction. Technical report.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343.