Volatility Modeling: From First Principles to The Mixture Models

Tashfeen Omran

January 20, 2026

Contents

1 Introduction

I stumbled on some posts made by a statistics PhD on Linkedin that were full of comments of practitioners from Morgan Stanley and other firms, knowing that most research papers are garbage, I thought I can learn several things from their responses and discussions.

2 Conceptual Foundations and Intuition

Before diving into complex modeling, we must establish the vocabulary that describes how real markets deviate from the theoretical assumptions of the Black-Scholes-Merton (BSM) model.

2.1 Market Dynamics: Skewness and Kurtosis

Standard models often assume a Normal (Gaussian) distribution of log-returns. However, market reality is different:

- **"The Elevator Down":** This phrase describes **Negative Skewness**. Markets tend to drift upwards slowly ("taking the stairs") but crash downwards rapidly ("taking the elevator"). BSM assumes symmetry; reality is asymmetric.
- **Fat Tails (Leptokurtosis):** Real markets exhibit extreme events far more frequently than a standard Bell Curve predicts. This is known as high kurtosis.

2.2 The Price-Volatility Relationship

In equity markets, the correlation between Spot Price (S) and Volatility (σ) is typically negative (Price $\downarrow$, Vol $\uparrow$). However, we discussed specific "broken" regimes where Price $\uparrow$ and Vol $\uparrow$:

1. **The Dot-Com Bubble (1999):** Prices rose, but fear of a crash caused demand for Puts to rise, pushing Implied Volatility (IV) up simultaneously.
2. **GameStop (Jan 2021) - The Short Gamma Trap:**
   - Market Makers sold Calls (becoming Short Call $\rightarrow$ Short Gamma).
   - As prices rose, the Delta of those Calls increased.
   - To hedge, Market Makers had to buy more stock, driving prices higher.
   - This chaotic feedback loop caused Volatility to skyrocket alongside Price.
3. **Earnings Announcements:** Prices may drift up slightly, but IV crushes upwards due to the binary uncertainty of the news event.

2.3 Probability vs. PDF

A key distinction clarified in our discussion:

- **PDF (Probability Density Function):** The curve $f(x)$. It represents the relative likelihood of a value.
- **Probability:** The area under the curve, calculated via the definite integral:

$$P(A \leq S_T \leq B) = \int_A^B f(S_T)\, dS_T$$

3 The Calculus of Probability: The Breeden-Litzenberger Theorem

This theorem is the "Holy Grail" that connects Option Prices to Probability Densities. It allows us to extract the market's implied probability distribution directly from the option chain.

3.1 Statement of the Theorem

The Risk-Neutral Probability Density Function (PDF) of the stock price at maturity T is proportional to the second derivative of the Call Price C with respect to the Strike K:

$$f_S(K) = e^{rT}\frac{\partial^2 C(K)}{\partial K^2}$$

Intuition: The "curvature" of the option prices reveals where the market thinks the stock will land.

3.2 Step-by-Step Derivation

We derive this from First Principles, addressing the specific confusion regarding Leibniz's Rule.

Step 1: Price as an Expectation

The price of a European Call under the Risk-Neutral measure $\mathbb{Q}$ is the discounted expected payoff:

$$C(K) = e^{-rT}\int_{-\infty}^{\infty} \max(S_T - K, 0) f_S(S_T)\, dS_T$$

Since the payoff is zero when $S_T < K$, the lower bound becomes K:

$$C(K) = e^{-rT}\int_{K}^{\infty} (S_T - K) f_S(S_T)\, dS_T$$

Step 2: The First Partial Derivative (Leibniz's Rule)

We differentiate with respect to K. Note that K appears in two places: inside the integrand $(S_T - K)$ and as the lower limit of the integral.

$$\frac{\partial C}{\partial K} = e^{-rT}\frac{\partial}{\partial K}\left(\int_{K}^{\infty} (S_T - K) f_S(S_T)\, dS_T\right)$$

Using Leibniz's Rule:

$$\frac{\partial}{\partial K}\int_{a(K)}^{b(K)} g(S_T, K)\, dS_T = \int_a^b \frac{\partial g}{\partial K}\, dS_T + g(b, K)\cdot b'(K) - g(a, K)\cdot a'(K)$$

Here, $a(K) = K$.

1. **Inside Derivative:** $\frac{\partial}{\partial K}(S_T - K) = -1$.
2. **Boundary Derivative:** When evaluating the boundary term at the lower limit $S_T = K$, the term $(S_T - K)$ becomes $(K - K) = 0$. *This explains why the boundary term disappears in this step.*

Result:

$$\frac{\partial C}{\partial K} = -e^{-rT}\int_{K}^{\infty} f_S(S_T)\, dS_T$$

This represents the negative cumulative probability (roughly $-$Delta).

Step 3: The Second Partial Derivative

We differentiate again with respect to K:

$$\frac{\partial^2 C}{\partial K^2} = -e^{-rT}\frac{\partial}{\partial K}\int_{K}^{\infty} f_S(S_T)\, dS_T$$

Using the Fundamental Theorem of Calculus for the lower bound: $\frac{d}{dx}\int_x^b f(t)dt = -f(x)$. The derivative of the integral introduces a negative sign, which cancels the existing negative sign:

$$\frac{\partial^2 C}{\partial K^2} = -e^{-rT}\left(-f_S(K)\right) = e^{-rT} f_S(K)$$

4 The Proposed Solution: Mixture Dynamics

The Brigo & Mercurio paper proposes that since a single BSM density fits poorly, we should use a weighted sum of densities.

4.1 The Gaussian Mixture Model (GMM)

The density $p(S)$ is defined as:

$$p(S) = \sum_{i=1}^{n} \lambda_i p_i(S) \quad \text{where} \quad \sum \lambda_i = 1$$

For a 2-regime model:

$$p(S) = \lambda \cdot p_{\text{Quiet}}(S; \sigma_1) + (1 - \lambda) \cdot p_{\text{Crash}}(S; \sigma_2)$$

This allows the model to capture the "Smile" by mixing a low-volatility curve with a high-volatility curve.

4.2 Crucial Distinction: Sums vs. Mixtures

A specific point of confusion addressed in our chat:

- **Sum of Variables ($X + Y$):** By the Central Limit Theorem, adding random variables tends toward a Normal distribution.
- **Mixture of Distributions:** Flipping a coin to choose between Distribution A and Distribution B.
- **Visual:** If you mix a distribution centered at 50 and one centered at 100, you get a **Bimodal** ("Camel Hump") distribution, NOT a single Normal distribution.

5 Critiques and The Reality Gap (The Debate)

Despite the mathematical elegance of Mixture Models, practitioners argue against them for specific reasons.

5.1 Extrapolation: "The Wings Fall Off"

This refers to the Deep Out-of-the-Money (OTM) tails.

- **Gaussian Decay:** The Lognormal distribution decays proportional to e^{-x^2}. For extreme events (e.g., $x = 10$), this value is effectively zero.
- **Power Laws (Reality):** Real markets follow Power Laws (Pareto distributions) which decay proportional to $x^{-\alpha}$ (e.g., $1/x^3$).
- **The Consequence:** A mixture of Gaussians is still Gaussian in the deep tail. It will underprice extreme crash risk compared to a Power Law.

5.2 Interpolation: "Spikiness" and Overfitting

- **"Knobs":** The model has many parameters (λ_i, σ_i). These are "knobs" we can turn to force the model to hit market prices.
- **Overfitting:** If we use too many knobs, the resulting PDF can become "wobbly" or "spiky" (like a rollercoaster) to fit the noise in the data.
- **Unstable Greeks:** A wobbly PDF means the Delta and Gamma change unpredictably, making hedging a nightmare.

6 Advanced Arbitrage Concepts

6.1 Calendar Arbitrage

Arbitrage is not just about price; it is about consistency across time.

- **Variance Additivity:** Uncertainty generally grows with time. Total Variance is $V(T) = \sigma^2 T$.
- **The Law:** $V(T_2) > V(T_1)$ for $T_2 > T_1$.
- **The Failure Mode:** If we calibrate the Mixture Model "slice-by-slice" (independently for each month), we might find that the 1-month crash risk is so high that $V(T_1) > V(T_2)$. This implies **Negative Forward Variance**, which is physically impossible and represents an arbitrage opportunity.

7 The Industry Standard Landscape

7.1 Stochastic Volatility Models

Instead of mixing static densities, these models assume Volatility itself is a random process.

- **Heston Model:** Assumes volatility is mean-reverting (pulls back to a long-term average θ).
- **SABR Model:** "Stochastic Alpha, Beta, Rho." Widely used in Rates/FX.
  - The β parameter handles the backbone (Normal vs. Lognormal).
  - The formula includes a "Normalization Factor" $\frac{\alpha}{(FK)^{(1-\beta)/2}}$ to average the Strike and Forward price.

7.2 Terminology Check

- **Wiener Process (W_t):** Synonymous with Brownian Motion. $dW_t \sim N(0, dt)$.
- **Gamma:** $\frac{\partial^2 C}{\partial S^2}$ (Convexity with respect to Spot).
- **Butterfly/Density:** $\frac{\partial^2 C}{\partial K^2}$ (Convexity with respect to Strike).

8 Synthesis: The Extremistan Hierarchy

We concluded with a breakdown of the "Extremistan" meme, illustrating the levels of modeling sophistication:

1. **Level 1 (Beginner):** Uses Mean and Standard Deviation. Fails because "Bill Gates walking into a bar" skews the average. (Assumes Mediocristan).
2. **Level 2 (Intermediate):** Rejects models because "Wealth follows a Power Law!" (Recognizes Fat Tails).
3. **Level 3 (Practitioner):** "Actually, Lognormal fits the bulk of the data better." (Pragmatism).
4. **Level 4 (Expert):** "A High-Volatility Lognormal mimics a Power Law."
   - *Insight:* We don't necessarily need new math. By using Mixture Models or Stochastic Volatility to inject a regime of massive σ, the Gaussian tail flattens out enough to behave like a Power Law for all practical purposes.

References

[1] Brigo, D., Mauri, G., & Mercurio, F. (2002). Lognormal-mixture dynamics and calibration to volatility smiles and skews. *International Journal of Theoretical and Applied Finance*, 5(04), 427-446.

[2] Brigo, D., & Mercurio, F. (2006). *Interest Rate Models-Theory and Practice: With Smile, Inflation and Credit*. Springer Science & Business Media.

[3] Breeden, D. T., & Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *Journal of Business*, 621-651.