Dynamic Volatility Modeling for Option Hedging

A Comparative Backtest of HV, GARCH, and GJR-GARCH Models

Dhanashree Somani

The Erdös Institute
Quant Finance Bootcamp 2025

November 7, 2025

Introduction & Objective

- The standard Black-Scholes (BS) model's biggest weakness is its assumption of **constant volatility**.
- Real-world volatility is dynamic, possibly exhibiting:
  - **Clustering:** High-volatility periods follow high-volatility, low follows low.
  - **Leverage Effect:** Volatility increases more when prices fall.
- **Objective:** To test if dynamic volatility models (GARCH & GJR-GARCH) provide more effective delta-hedging performance than a simple Historical Volatility (HV) model.

The Models: Three Contenders

1. **30-Day Historical Vol (HV)**
   - *Benchmark.* Simple, adaptive, but naive. Ignores clustering and leverage effect.
2. **GARCH(1,1)**
   - *Symmetric Model.* Captures **volatility clustering**.
   - Treats +5% and -5% shocks identically.
3. **GJR-GARCH(1,1)**
   - *Asymmetric Model.* Captures clustering *and* the **leverage effect** via the gamma (γ) parameter.

Hypothesis: The "smarter" GJR-GARCH model should be the best performer.

Methodology: Rolling-Window Backtest

- **Assets:** AAPL, GOOGL, MSFT, NVDA, TSLA
- **Simulation Loop (Run Daily):**
  - "Sell" a new 30-day at-the-money (ATM) call option.
  - Forecast volatility using all 3 models (500-day rolling window for GARCH).
  - Calculate BS Delta using each of the 3 different volatility forecasts.
  - Re-hedge this position daily.
- **Key Metric: Variance of P&L (Hedge Error)**
  - A perfect hedge has a P&L variance of 0.
  - The model with the **lowest variance** is the most stable and effective hedge.

Results 1: The "Textbook Case" (GOOGL & MSFT)

Finding: For GOOGL and MSFT, the hypothesis was **correct**. The asymmetric GJR-GARCH model was the clear champion.

Ticker *vs. HV*	**GARCH(1,1)** *Var. Reduction*	**GJR-GARCH** *Var. Reduction*
GOOGL	18.30%	**20.05%**
MSFT	14.53%	**18.49%**

Table: Variance Reduction vs. 30-Day Historical Vol. GJR-GARCH successfully reduced risk.

Results 2: The "Whipsaw Case" (AAPL & NVDA)

Finding: For AAPL and NVDA, the "smarter" GJR-GARCH model **failed**, performing worse than the simpler GARCH(1,1).

Ticker *vs. HV*	**GARCH(1,1)** *Var. Reduction*	**GJR-GARCH** *Var. Reduction*
AAPL	-4.06%	**-12.65%**
NVDA	**12.48%**	7.21%

Table: Variance Reduction vs. 30-Day Historical Vol. GJR-GARCH *increased* risk for AAPL.

Results 3: The "Overfitting Case" (TSLA)

Finding: For TSLA, the symmetric GARCH(1,1) was the clear winner. The GJR-GARCH model failed to provide a better hedge.

Ticker *vs. HV*	**GARCH(1,1)** *Var. Reduction*	**GJR-GARCH** *Var. Reduction*
TSLA	**11.73%**	10.43%

Table: Variance Reduction vs. 30-Day Historical Vol. The simpler model won.

Analysis: The GJR Gamma (γ) Parameter

The γ **(gamma)** parameter measures the leverage effect. A high, significant γ explains the results.

Ticker	Gamma (γ)	P-value	Finding
Group 1			
GOOGL	0.0477	0.000	**Moderate Effect**
MSFT	0.0989	0.008	**Moderate Effect**
Group 2			
AAPL	**0.1447**	0.000	**Extreme Effect**
NVDA	**0.1417**	0.022	**Extreme Effect**
Group 3			
TSLA	-0.0046	*0.706*	**Not Significant**

Table: GJR-GARCH(1,1) Parameter Fits (2017-2024 data)

Grand Conclusion: "No One Size Fits All"

- **Case 1 (GOOGL, MSFT):** GJR-GARCH works perfectly. Their *moderate* and *significant* leverage effect is modeled well.
- **Case 2 (AAPL, NVDA):** GJR-GARCH fails. Their *extreme* leverage effect (γ is 1.5x-3x larger) creates a "whipsaw" when combined with a naive BS-delta. The model is **too reactive**.
- **Case 3 (TSLA):** GJR-GARCH fails. Its *gamma is not statistically significant* (p=0.7). The model is **over-parameterized** for an asset with no leverage effect.

Final Takeaway

A successful hedging strategy requires a model that is **correctly specified** for the unique volatility structure of the asset.

Thankyou!