Hypothesis Testing of Standard Assumptions

Erdös Quantitative Finance – Mini Project 2

Guoqin Liu

June 26, 2025

Motivation & Research Questions

- Text-book models often assume log-returns $\sim \mathcal{N}(\mu, \sigma^2)$.
- **Questions**
  1. Are there full periods where this holds for major ETFs?
  2. Do *rolling* 1-year windows ever look normal?
  3. Does trimming 1% tails help?
  4. Does an equal-weight portfolio of 5 assets look normal?

Data & Methodology

- **Assets**: SPY, QQQ, IWM (equities), TLT (US Treasury), GLD (gold)
- **Sample**: 2010-01-01 – June 26, 2025 (about 3 900 daily obs.)
- **Tests** (all four must keep H_0 to "Accept"):
  1. Shapiro–Wilk (skew + kurtosis)
  2. Anderson–Darling (EDF-based)
  3. Jarque–Bera (skew/kurt excess)
  4. Kolmogorov–Smirnov (CDF distance)
- **Extras**: 252-day rolling windows, 1 % Winsorising, equal-weight portfolio.

Whole-Period Results (2010-June 26, 2025)

Ticker	Shapiro p	AD p	JB p	KS p	Accept?
SPY	≈ 0	0.15	≈ 0	≈ 0	No
QQQ	≈ 0	0.15	≈ 0	≈ 0	No
IWM	≈ 0	0.15	≈ 0	≈ 0	No
TLT	≈ 0	0.15	≈ 0	5×10^{-4}	No
GLD	≈ 0	0.15	≈ 0	≈ 0	No

- With $n \approx 3\,900$, even slight deviations are highly significant.
- Heavy tails and volatility clustering drive the rejection.

Rolling 252-Day Windows

- Scan every 1-year window for each ticker.
- **SPY**: 1–2 % of windows accept normality (e.g. Jun 2014 – Jun 2015 period of low vol).
- Other tickers show similar or lower pass rates; many have zero.

Effect of Trimming 1 % Tails

- Winsorising reduces excess kurtosis from $\sim$8–10 to $\sim$4–5.
- Shapiro & JB p-values often rise above 0.05, **but** AD/KS still reject.
- Conclusion: heavy-tail mass extends beyond the outermost 1 %.

Equal-Weight Portfolio of 5 ETFs

- Mix: SPY, QQQ, IWM, TLT, GLD (weights = 20% each).

Shapiro	AD	JB	KS	Accept?
2.6×10^{-40}	0.15	0	4.3×10^{-17}	No

- Diversification insufficient: correlated tail events persist.
- Volatility clustering survives aggregation.

Key Takeaways

1. Daily log-returns of major ETFs are not normal over long spans.
2. Short calm regimes *can* look Gaussian.
3. Light tail trimming (1 %) rarely enough.
4. Small diversified portfolios still inherit fat tails.

Fixes: lower frequency, volatility scaling, heavy-tailed distributions, or explicit GARCH modelling.

Next Steps

- **Volatility-scaled returns**: test $r_t/\hat{\sigma}_t$ for normality.
- Compare Normal vs Student-t vs Skew-t fits (AIC/LLH).
- Stress-testing VaR/ES under non-Gaussian assumptions.
- Extend to global assets and alternative risk-parity weights.