📘 Quantitative Finance Regression Analysis with Alpha Signals

This notebook goes beyond just cross-sectional regressions. It demonstrates a complete, end-to-end quantitative finance analysis pipeline using real stock market and macroeconomic data. Below is a complete summary of what was implemented, along with simplified explanations and key takeaways.

📌 Project Overview

This project analyzes a set of U.S. stocks (mostly FAANG and major indices) using:

- **Time-series regression**: To understand how each factor affects a specific stock over time.
- **Cross-sectional regression**: To evaluate which factors consistently explain differences between stocks in the same week.

The goal is to identify **statistically significant alpha signals** and understand how macro factors (like VIX, interest rates, dollar strength) impact returns.

🔧 Data Preparation

- **Stocks Analyzed**: AAPL, MSFT, AMZN, GOOGL, META, NFLX, SPY, QQQ, IWM
- **Macroeconomic Proxies**:
  - VIX (market volatility)
  - TLT (interest rate proxy)
  - DXY (U.S. dollar strength)
- **Sources**: All data collected using `yfinance` from Yahoo Finance.

🧠 Alpha Signal Engineering

Signal	Description
`Mom1M`	1-month price momentum (short-term)
`Mom3M`	3-month momentum (medium-term)
`Vol_Surp`	Volume surprise: how different today's volume is from 20-day average
`Volatility`	21-day rolling standard deviation of returns
`Beta`	Rolling 60-day beta with SPY
`VIX_chg`	Daily percentage change in VIX (fear index)
`TLT_ret`	Daily return of TLT (20Y bond ETF)
`DXY_chg`	Daily percentage change in USD index

All signals are aligned and merged per (Date, Ticker) for predictive modeling.

📈 Time-Series Regression (AAPL Case Study)

CAPM Regression (AAPL ~ SPY)

- AAPL's beta ≈ 1.2 → AAPL moves 20% more than the market on average
- R^2 ≈ 30-50% → reasonable explanatory power for a single factor

Multivariate Regression

- Factors: SPY, Mom1M, Mom3M, Vol_Surp, Volatility, Beta, VIX, TLT, DXY
- Coefficients interpreted in financial context (e.g. negative VIX → rising fear hurts AAPL)
- R^2 improved to ~40%+ with all factors included

Diagnostics

- Residuals were normally distributed (QQ plot)
- No strong autocorrelation or heteroskedasticity
- Inference from t-stats and p-values reliable

🧪 Out-of-Sample Testing

- Training: 2018–2021, Testing: 2022–2024
- Out-of-sample R^2 lower (as expected), but core macro effects (VIX, SPY) remained strong

- Demonstrated importance of **avoiding overfitting** and checking for signal robustness

---

🤝 Cross-Sectional Regression

- Weekly regression across 9 stocks
- Factors: `Mom1M`, `Vol_Surp`, `Volatility`, `Beta`

Key Results:

Factor	Avg Coeff	t-Stat	Interpretation
Mom1M	-0.15	-2.5	Reversal effect (recent winners underperform)
Vol_Surp	+0.05	+1.2	Not significant, some hint of effect
Volatility	-0.08	-2.0	Significant low-volatility premium
Beta	-0.02	-0.5	No significant premium for high-beta

📌 Summary of Findings

- **Market returns (SPY)** were the most powerful explanatory factor.
- **VIX spikes** consistently led to **stock price drops** (strong negative effect).
- **Short-term reversal**: 1-month momentum had a negative effect → contrarian strategy.
- **Medium-term momentum (3M)** had a mild positive effect → continuation strategy.
- **Low-volatility stocks** outperformed more volatile ones, aligning with the low-vol anomaly.
- **Volume surprise** was inconsistent.
- **Beta** did not produce reliable excess returns.

🧠 Practical Use Cases

- **Alpha generation**: Form long-short portfolios based on momentum or volatility.
- **Risk control**: Use VIX, TLT, DXY exposure to manage macro risk.
- **Performance attribution**: Use model coefficients to decompose return sources.

🧩 Next Steps & Extensions

- Expand universe to include 100–500 stocks
- Incorporate Fama-French style factors (Size, Value)
- Use ridge regression or machine learning for nonlinear signals
- Backtest factor-ranked portfolios

🏁 Conclusion

This notebook represents a complete quantitative research pipeline:

- 📦 From raw data → clean features
- 📉 From regression modeling → interpretable coefficients
- 📊 From statistical inference → financial strategy