TSA project presentation

Giacomo Fantato · **K-16958**
Federico Mennella · **K-17572**
Domenico Castaldo · **K-17409**

Exploratory Data Analysis

EDA is performed on the training set and test set

Data preparation

Ticker chosen: Equities (Boeing, traded on NYSE) | Equity Index (S&P 500) | Commodity (Gold) | Currency (USD/PLN) | Cryptocurrency (Ethereum)

Data import: we use the *data_engineering.py* class to import data directly from yahoo Finance. A kill switch function, automatically imports data from local CSVs in case yahoo finance is not available

Clean missing data: remove rows with NaN values for consistency.

Dataset alignment: align datasets to the calendar that has fewer days (stocks that follow the NYSE calendar)

Splitting the datasets: following the design specifications we split in-sample (train set) and out-of-sample (test set). the initial analyses were performed on both sets, from modeling onwards only on the train.

Train set: training the best GARCH family model

VaR forecasting

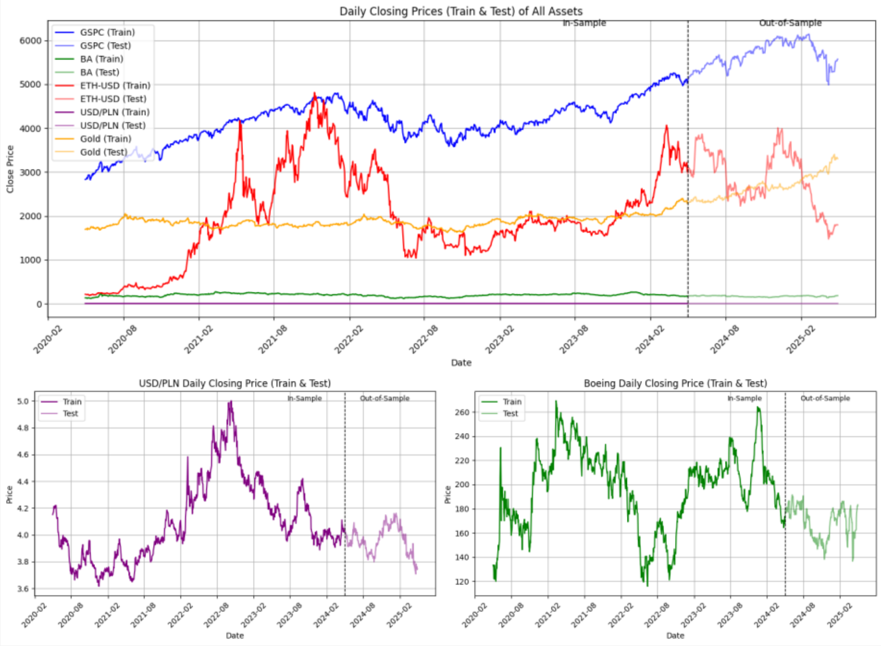
Train (5 or 10 folders) | Validation

Found best model for each index

Test

NB: The entire portfolio is always balanced in 20% for each ticker!

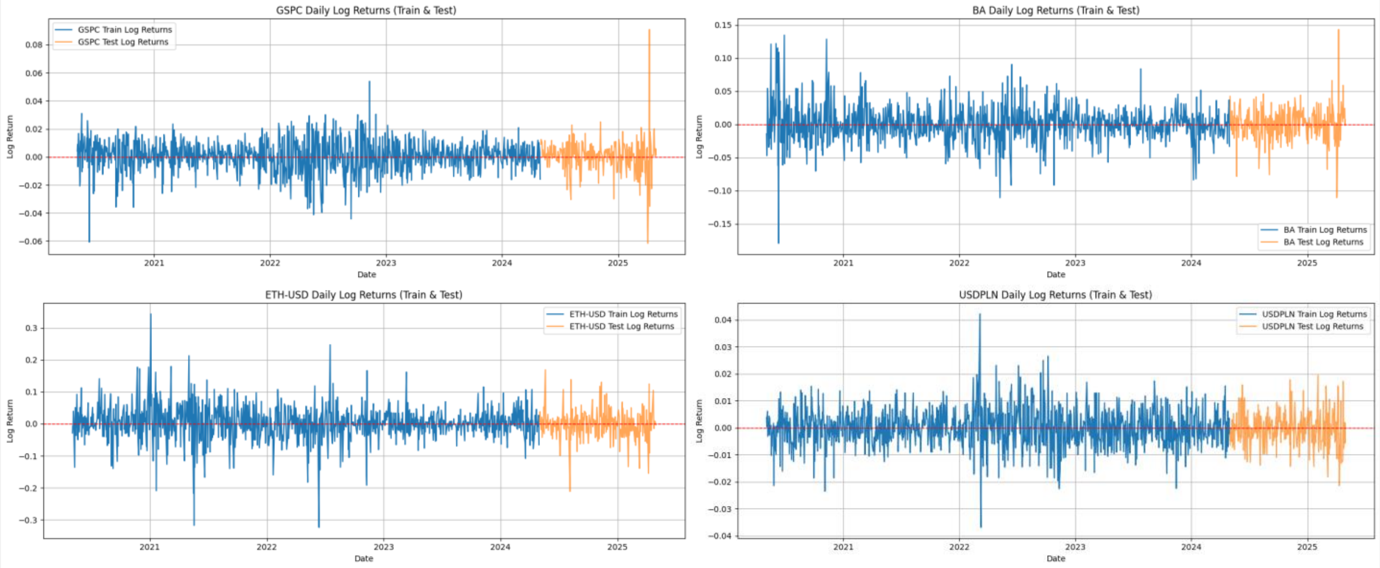
Data exploration – *closing prices*



A black dashed vertical line indicates the threshold between the training and test periods. The high variability of Ethereum compared to the other assets highlights its typically more speculative and volatile nature.

The chart displays the daily closing prices and upcoming volatility of each asset solely for informational purposes, without implying any investment advice or forecasts.

Data exploration – *volatility*



Train set | Test set

Returns stylized facts

Investigating the **training set** to identify key features of financial time series behavior

Investigating the possible leptokurtosis

Means (daily) returns are generally very close to zero, with Ethereum (ETH-USD) showing the highest average daily return. USD/PLN is the only asset with a slightly negative mean.

Skewness indicates asymmetry in return distributions. S&P 500, Gold, and the Portfolio exhibit negative skewness, suggesting a higher probability of extreme negative returns.

Kurtosis values are significantly greater than 3 for all assets, confirming the presence of fat tails (leptokurtosis) and extreme events. ETH and Boeing display the most pronounced excess kurtosis.

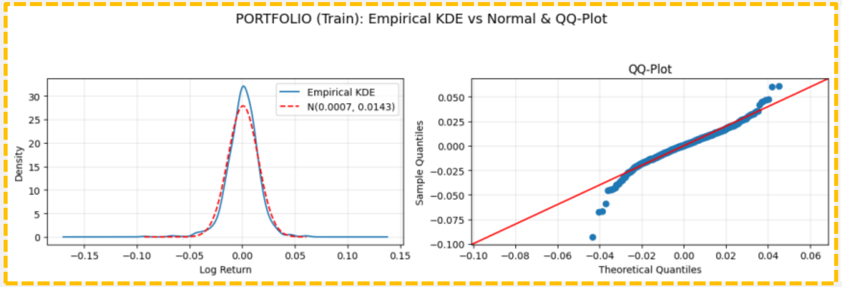
The Jarque-Bera test strongly rejects normality for all return series (p-value < 0.05), confirming that asset returns are not normally distributed.

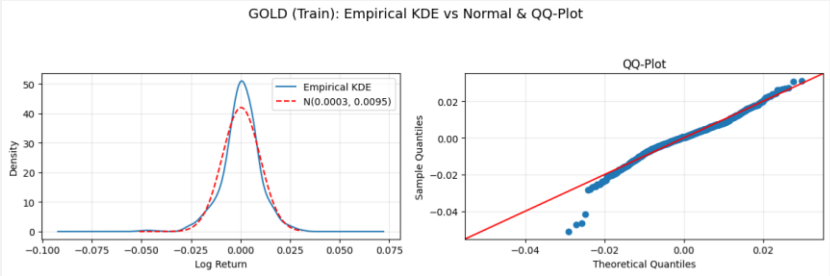
The Ljung-Box test on raw returns reveals significant autocorrelation in USD/PLN/BA and the Portfolio, while other assets do not show serial dependence.

These findings confirm several classic stylized facts: non-normality, fat tails, skewness, and autocorrelation, providing strong motivation for using GARCH-type models to capture volatility dynamics.

NB: *all results are based on daily log returns in the training period.*

Distributional assessment via KDE and QQ-Plots



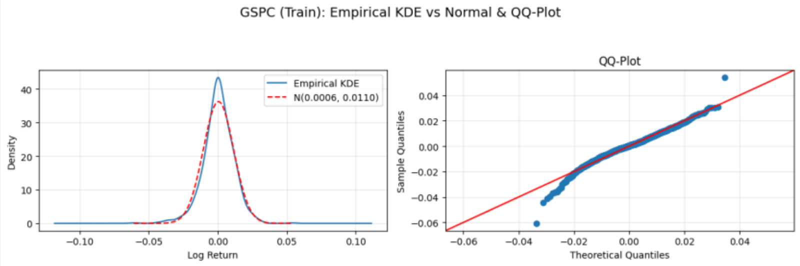


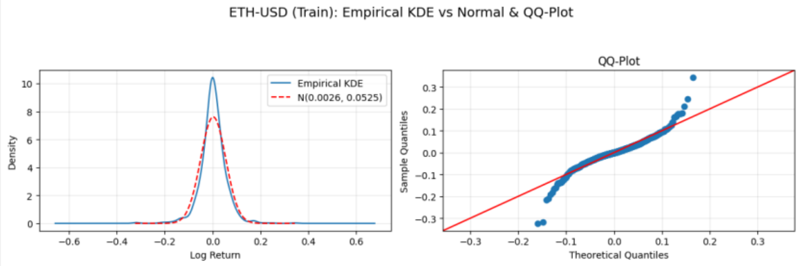
KDE reveals a distribution that is more peaked and has heavier tails than the fitted normal distribution (red dashed line) → **leptokurtosis**.

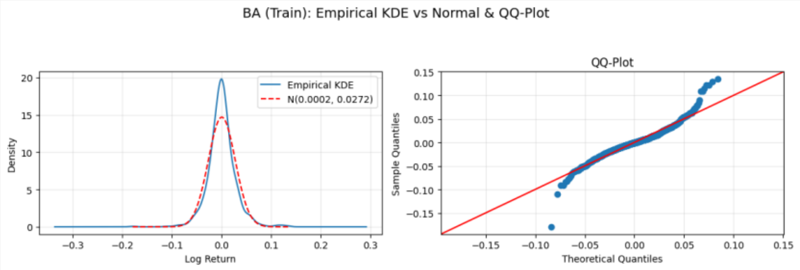
QQ-Plot further confirms departure from normality, especially in the tails, where extreme quantiles deviate significantly from the 45° reference line.

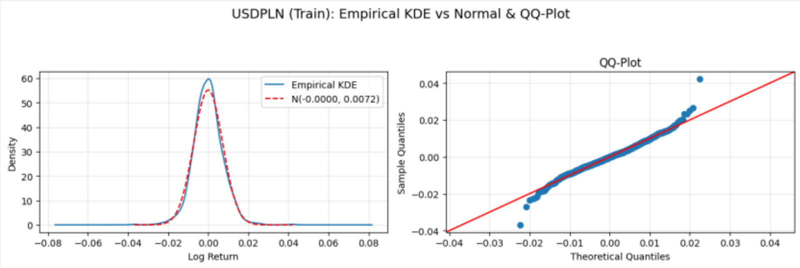
These results support previous findings (e.g., Jarque-Bera test) that portfolio returns are not normally distributed, with fat tails and excess kurtosis.

Distributional assessment via KDE and QQ-Plots



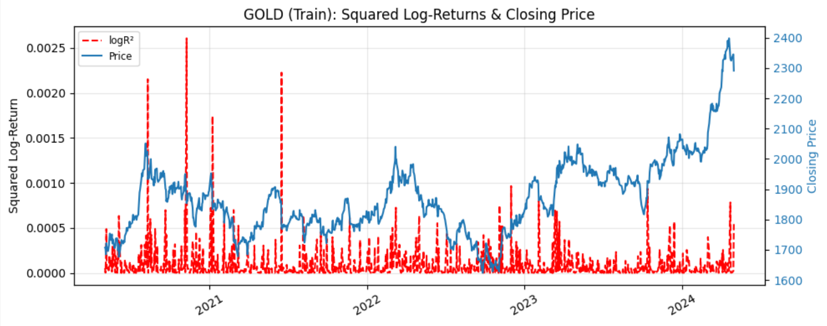


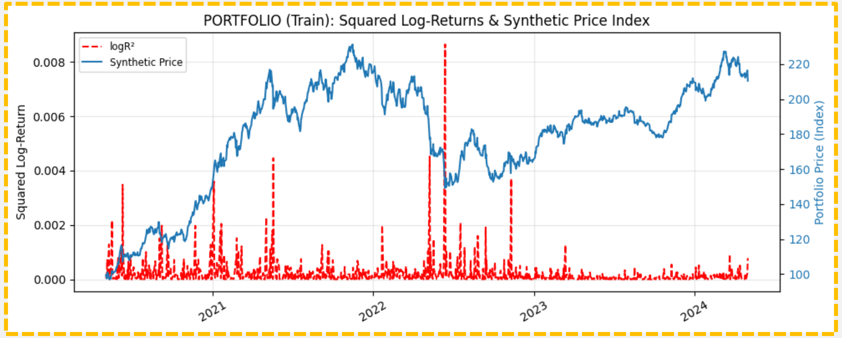




The portfolio's return distribution deviates from normality, reinforcing the need for volatility models that accommodate heavy tails, such as GARCH-type models.

Investigating the volatility





The pattern of bursts of squared logical returns followed by periods of calm is a distinctive feature of volatility clustering.

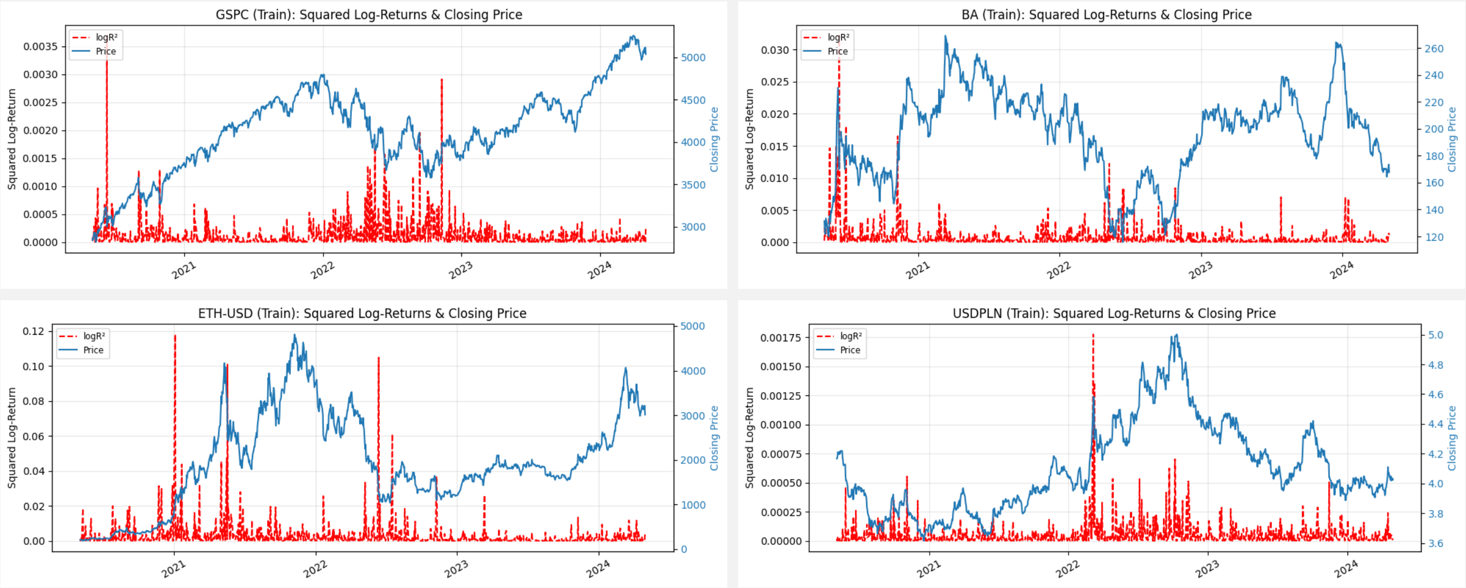
Clusters of red peaks reflect sustained high-volatility episodes, not isolated shocks.

Flat bands indicate periods of persistently low volatility.

This behavior implies positive autocorrelation in volatility: high (or low) volatility tends to persist over time.

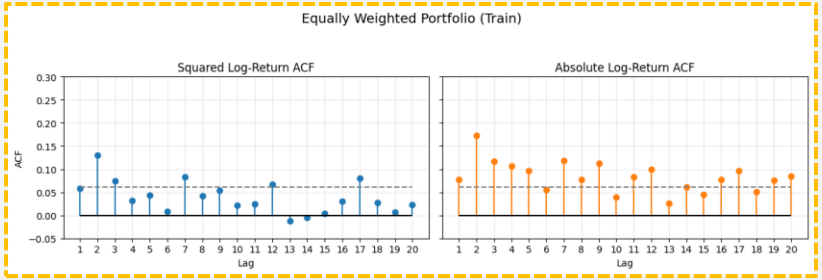
ARCH/GARCH models are specifically designed to capture this persistence in variance

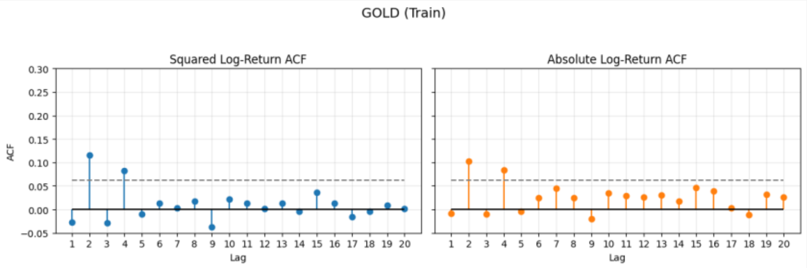
Investigating the volatility



To confirm clustering formally, we analyze the ACF of squared or absolute returns, both typically show slow decay, indicating long-memory in volatility.

ARCH effects among log-returns





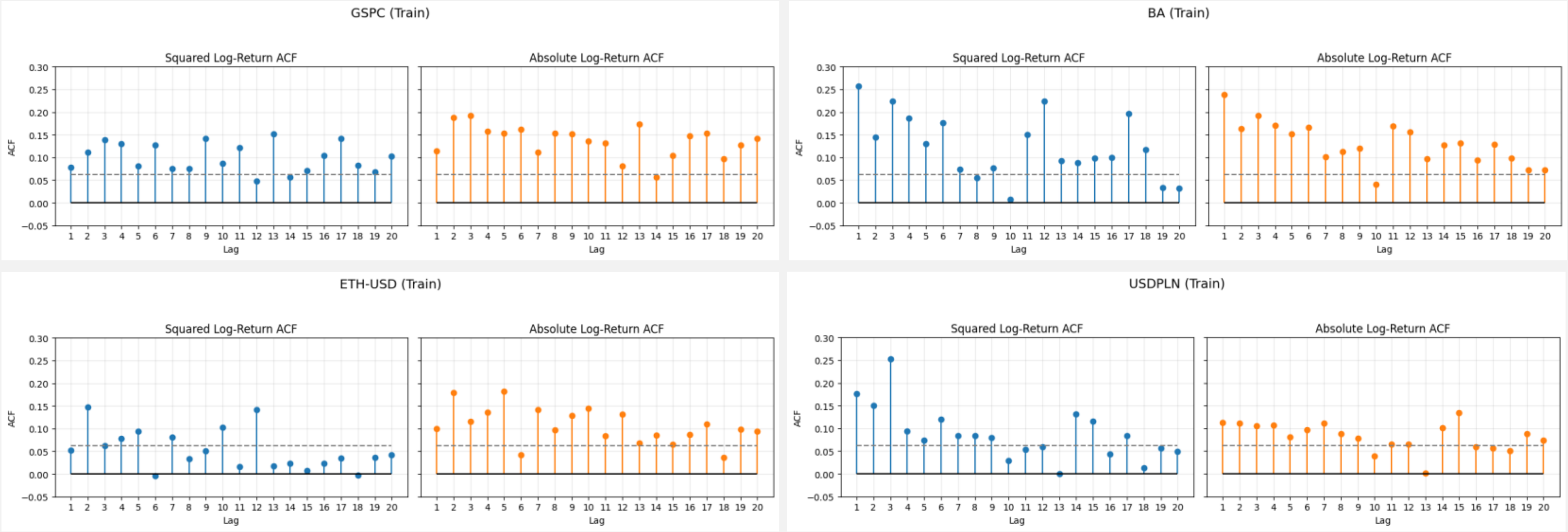
All four assets, S&P 500, Boeing, Ethereum, and USD/PLN, show *statistically significant autocorrelation* in both squared and absolute log-returns.

The presence of slowly decaying ACFs, particularly for absolute returns, highlights a persistent volatility structure over time. Even for USD/PLN, a relatively stable FX pair, volatility clustering is evident.

Boeing and Ethereum exhibit the strongest clustering, with sharp spikes in autocorrelation, suggesting highly volatile and reactive behavior.

The negative autocorrelation in Gold's (*safe-haven asset*) returns suggests possible **mean reversion** or **market microstructure effects**, such as *bid-ask bounce* or *short-term overreactions*.

ARCH effects among log-returns



This consistent autocorrelation structure across all assets supports the use of ARCH/GARCH models to capture time-varying volatility in financial returns. Gold seems not to have too many ARCH effects.

ARCH effects – LM testing

```
--- GSPC ---
LM Statistic: 45.294    p-value: 0.000
F Statistic:  9.432    p-value: 0.000

--- BA ---
LM Statistic: 113.083    p-value: 0.000
F Statistic:  25.347    p-value: 0.000

--- ETH-USD ---
LM Statistic: 34.527    p-value: 0.000
F Statistic:  7.109    p-value: 0.000

--- USDPLN ---
LM Statistic: 90.855    p-value: 0.000
F Statistic:  19.867    p-value: 0.000

--- GOLD ---
LM Statistic: 19.710    p-value: 0.001
F Statistic:  3.997    p-value: 0.001

--- Portfolio ---
LM Statistic: 24.310    p-value: 0.000
F Statistic:  4.953    p-value: 0.000
```

LM test

```
--- GSPC ---
ARCH Effects Present: YES

--- BA ---
ARCH Effects Present: YES

--- ETH-USD ---
ARCH Effects Present: YES

--- USDPLN ---
ARCH Effects Present: YES

--- GOLD ---
ARCH Effects Present: YES

--- Portfolio ---
ARCH Effects Present: YES
```

Our LM automated test

The Engle's LM-ARCH test confirms the presence of conditional heteroskedasticity in all in-sample log-return series, both individual assets and the portfolio.

All p-values are < 0.05, leading to **rejection of the null hypothesis of no ARCH effects**.

This provides strong statistical evidence of time-varying volatility, justifying the use of GARCH-family models.

The 'gold' absolute-return ACF likewise shows a couple of early significant bars. This pattern implies that gold volatility clusters only briefly: shocks have shorter lives, and the ARCH effect is far weaker here than in other assets

Additionally, our automated LM-ARCH testing module performs reliably and is *ready to be integrated into a fully automated modeling pipeline* (currently under development).

GARCH modelling

Let's evaluate the performance of our fully automated pipeline for selecting and estimating the optimal GARCH model.

Portfolio volatility approach

We are going to adopt two approaches in order to model the portfolio volatility, one of which is presented today

Direct approach

Involves modeling the univariate time series of the portfolio returns directly, applying standard volatility measures such as rolling standard deviation or univariate GARCH models. This approach treats the portfolio as a single asset.

Indirect approach

This method acknowledges that portfolio volatility depends both on the individual volatilities of the assets and on their covariances (co-volatilities). This can be done both in the univariate and multivariate case.

Automated GARCH model selection pipeline

This routine automates the selection of the best GARCH(p,q) model for each asset, fully aligned with the project's methodological requirements.

Features:

- Tests **multiple GARCH configurations**: (1,1), (2,1), (1,2), (2,2)
- Uses **rescaled log-returns** (×100) for numerical stability
- Applies strict diagnostic validation based on:

AIC minimization (for the choice) | **BIC** (only as info metric) | **Ljung–Box test** on residuals & squared residuals | **LM-ARCH** test on standardized residuals

Model acceptance criteria:

- Ljung–Box p-values > 0.05 (no autocorrelation)
- LM ARCH p-value > 0.05 (no ARCH effects)
- AIC is the lowest among all candidates

If no model meets all conditions, the best-AIC model is returned with warnings:

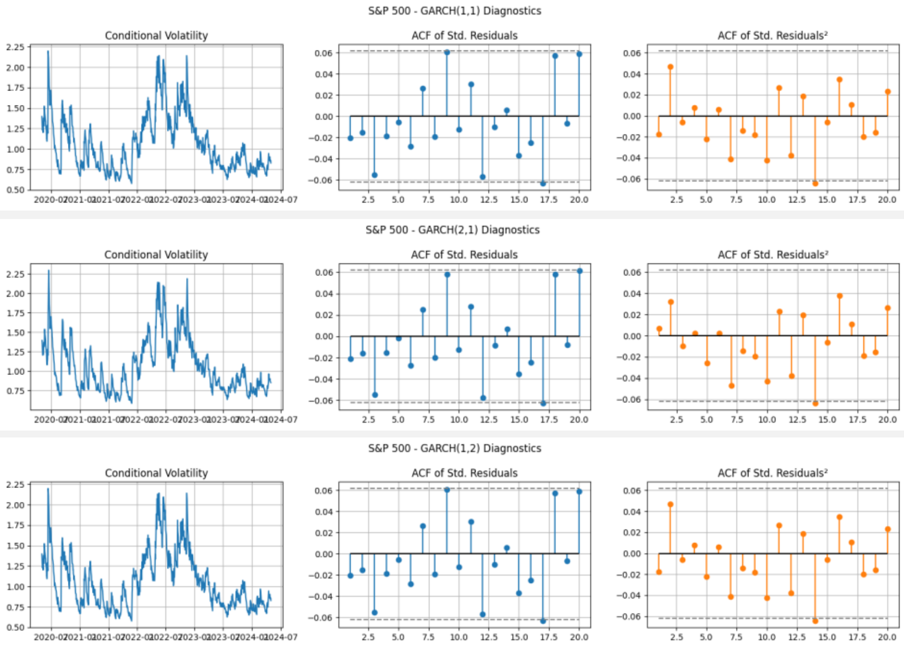
```
Best model for S&P 500 is GARCH(1,1) (AIC = 2863.96)
```

```
No model passed all diagnostics. Best by AIC for USD-PLN is GARCH(1,1) (AIC = 2116.31)
```

NB: *all the next models were estimated assuming (I) a constant mean and (II) Student's t-distributed residuals.*

Automated GARCH model selection pipeline

Example on S&P500 asset:




```
Best model for S&P 500 is GARCH(1,1) (AIC = 2863.96)

{'Model': 'GARCH(1,1)',
 'AIC': 2863.9574680262485,
 'BIC': 2888.5211821287144,
 'LB_resid_p': {5: 0.5332341137215522,
  10: 0.4475956569813382,
  15: 0.4042630548739195,
  20: 0.12386410605935147},
 'LB_sqres_p': {5: 0.676606414088706,
  10: 0.6981477696433678,
  15: 0.5166395376011227,
  20: 0.6743416992615154},
 'ARCH_pval': 0.6952136945184295,
 'res':                      Constant Mean - GARCH Model Results
 ==============================================================================
 Dep. Variable:                      Close   R-squared:                       0.000
 Mean Model:                 Constant Mean   Adj. R-squared:                  0.000
 Vol Model:                          GARCH   Log-Likelihood:               -1426.98
 Distribution:      Standardized Student's t   AIC:                          2863.96
 Method:                Maximum Likelihood   BIC:                          2888.52
                                             No. Observations:                1005
 Date:                    mer, giu 04 2025   Df Residuals:                    1004
 Time:                            19:03:33   Df Model:                           1
                                Mean Model
 ==========================================================================
                  coef    std err          t      P>|t|    95.0% Conf. Int.
 --------------------------------------------------------------------------
 mu             0.1040  2.779e-02      3.744  1.809e-04 [4.958e-02,  0.158]
                               Volatility Model
 ============================================================================
                  coef    std err          t      P>|t|       95.0% Conf. Int.
 ----------------------------------------------------------------------------
 omega          0.0226  1.156e-02      1.955  5.059e-02 [-5.770e-05,4.524e-02]
 alpha[1]       0.0928  2.782e-02      3.334  8.560e-04    [3.823e-02,  0.147]
 beta[1]        0.8901  3.149e-02     28.267 8.831e-176      [  0.828,  0.952]
```

NB: for every GARCH model (automated and manually) we use a “modified” GARCH with a t-Student distribution instated normal, to compensate leptokurtotic behavior of our time series

Automated GARCH – manual check

Let's check manually the model – GARCH(1,1) over **S&P500 asset**

```
gspc_garch11 = arch_model(y = 100*np.log(gspc_train['Close']).diff().dropna(), mean = 'constant', vol =
'GARCH', p = 1, q = 1, dist = 't')
gspc_garch11_result = gspc_garch11.fit()
```

```
                     Constant Mean - GARCH Model Results
==============================================================================
Dep. Variable:                      Close   R-squared:                       0.000
Mean Model:                 Constant Mean   Adj. R-squared:                  0.000
Vol Model:                          GARCH   Log-Likelihood:               -1426.98
Distribution:  Standardized Student's t   AIC:                           2863.96
Method:                Maximum Likelihood   BIC:                           2888.52
                                            No. Observations:                 1005
Date:                    mer, giu 04 2025   Df Residuals:                     1004
Time:                            19:04:07   Df Model:                            1
                                Mean Model
============================================================================
                 coef    std err          t      P>|t|    95.0% Conf. Int.
----------------------------------------------------------------------------
mu             0.1040  2.779e-02      3.744  1.809e-04 [4.958e-02,  0.158]
                              Volatility Model
==============================================================================
                 coef    std err          t      P>|t|       95.0% Conf. Int.
------------------------------------------------------------------------------
omega          0.0226  1.156e-02      1.955  5.059e-02 [-5.770e-05,4.524e-02]
alpha[1]       0.0928  2.782e-02      3.334  8.560e-04   [3.823e-02,  0.147]
beta[1]        0.8901  3.149e-02     28.267 8.831e-176      [  0.828,  0.952]
                              Distribution
========================================================================
                 coef    std err          t      P>|t|  95.0% Conf. Int.
------------------------------------------------------------------------
nu             8.8189      2.329      3.786  1.530e-04 [  4.254, 13.384]
========================================================================

Covariance estimator: robust
```

Parameter interpretation

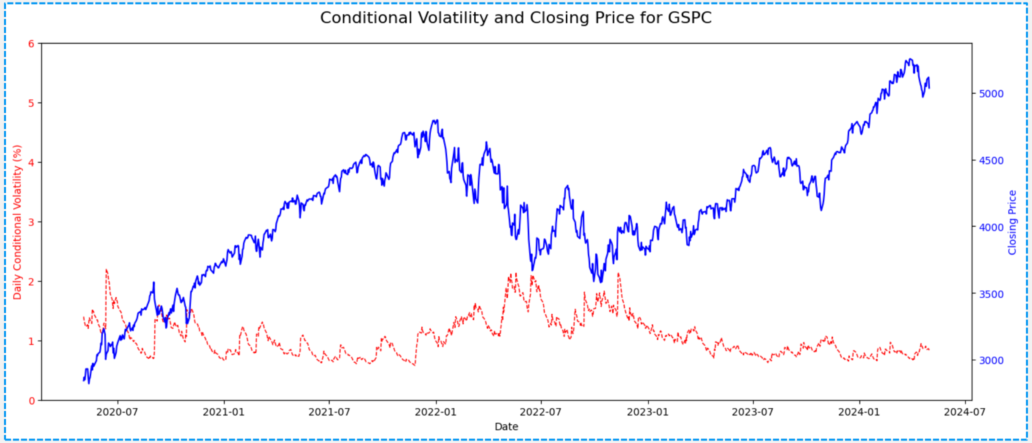
μ = 0.1040 → significant, positive mean return
ω = 0.0226 → baseline variance
α_1 = 0.0928 → reaction to past shocks
β_1 = 0.8901 → volatility persistence
$\alpha_1 + \beta_1$ = 0.9829 → stationarity condition satisfied
ν = 8.8191 → presence of fat tails (Student's t)

ARCH effects

The model has passed autocorrelation and ARCH effects tests on standardized residuals.

Automated GARCH – manual check

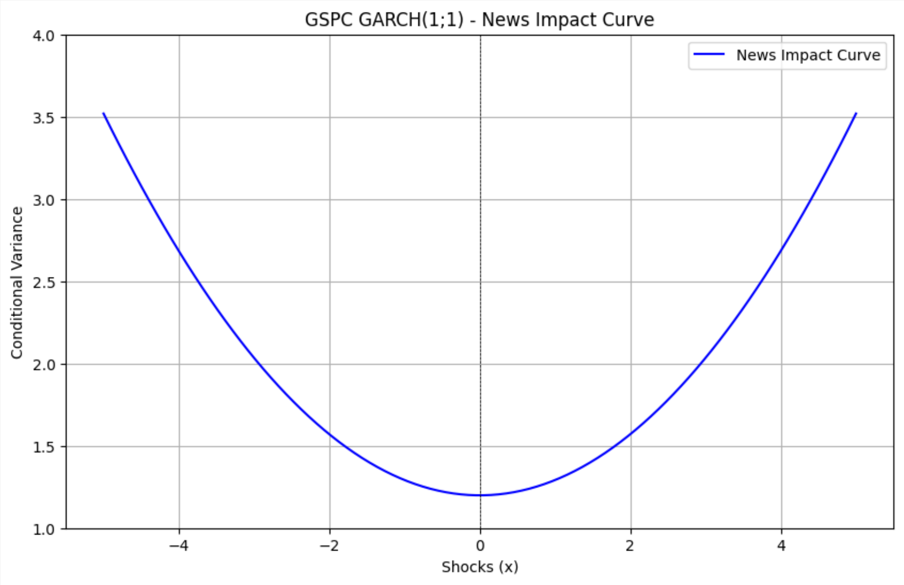
Let's check manually the model – GARCH(1,1) over **S&P500 asset**



This plot displays the **conditional volatility** (also known as the **conditional standard deviation**) estimated by the model. It shows significant spikes during periods of market turmoil, responding to new shocks in prices (and consequently in log-returns). The persistence of volatility is evident, as clusters of high (or low) volatility follow periods of elevated (or diminished) volatility.

Automated GARCH – manual check

Let's check manually the model – GARCH(1,1) over **S&P500 asset**



News impact curve

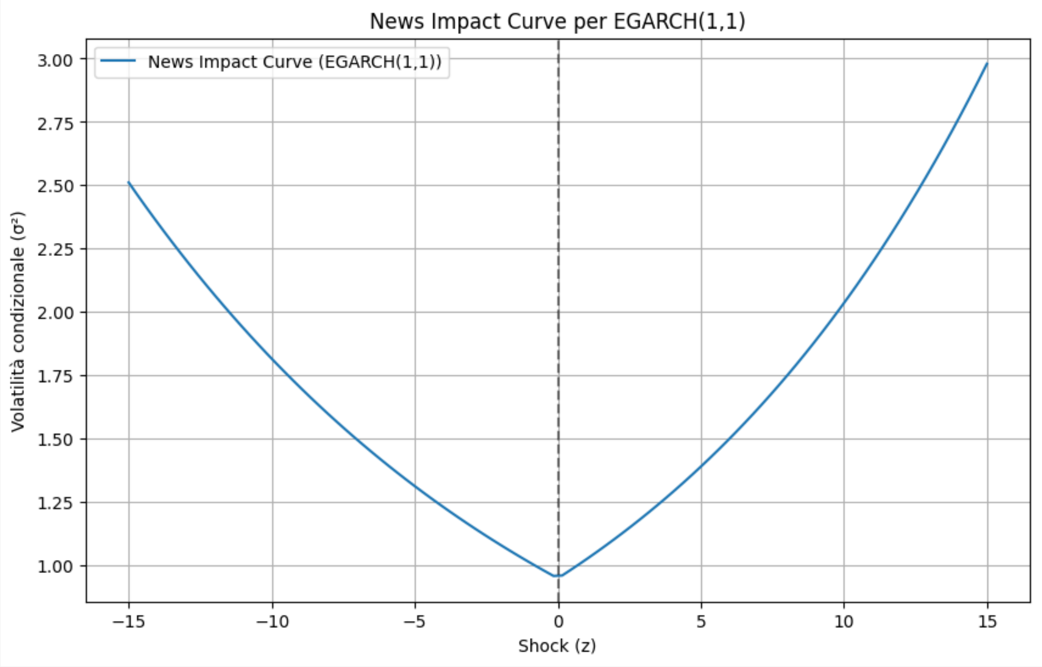
It shows the impact of shocks on the estimated conditional volatility, with the curve's shape indicating that both positive and negative shocks lead to an increase in volatility. As we know, the GARCH model does not account for the leverage effect: the relationship between volatility and shocks is modeled symmetrically.

Conclusions: our automated testing performs well, but result consistency should still be verified—manual review remains best practice for a limited number of time series.

Other GARCH-family models

Let’s check others GARCH-family model compatible with our time series

EGARCH – portfolio



Exploring EGARCH for Asymmetric Volatility

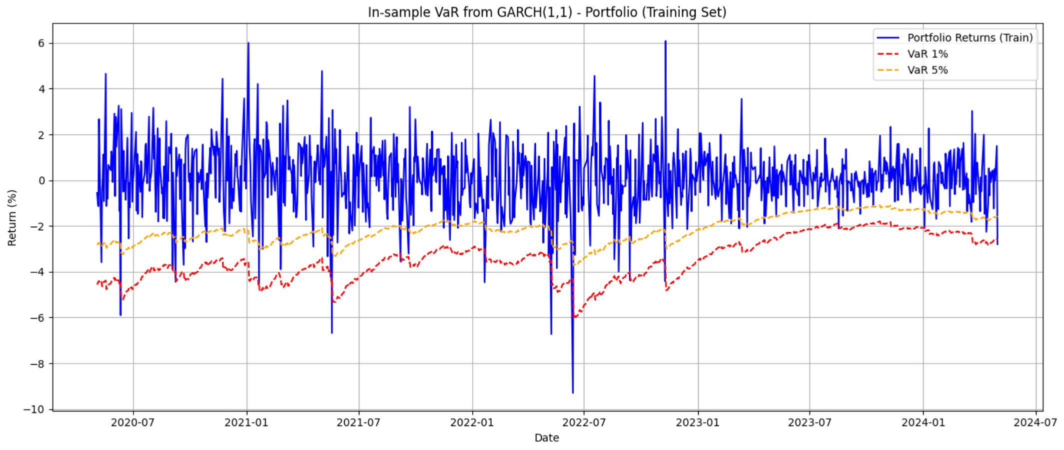
Unlike standard GARCH models, EGARCH captures volatility asymmetry—known as the *leverage effect*—where negative shocks tend to increase volatility more than positive ones of the same magnitude. This is evident in the News *Impact Curve* shown alongside.

The insignificance of γ suggests no strong evidence of asymmetric volatility in the portfolio. This is confirmed by the flat shape of the News Impact Curve (NIC) for the fitted EGARCH model.

Value at Risk - VaR

Performed on in-sample for the training | out-of-sample for the forecasting,
to *avoid data leakage*

VaR in-sample - GARCH(1,1) model

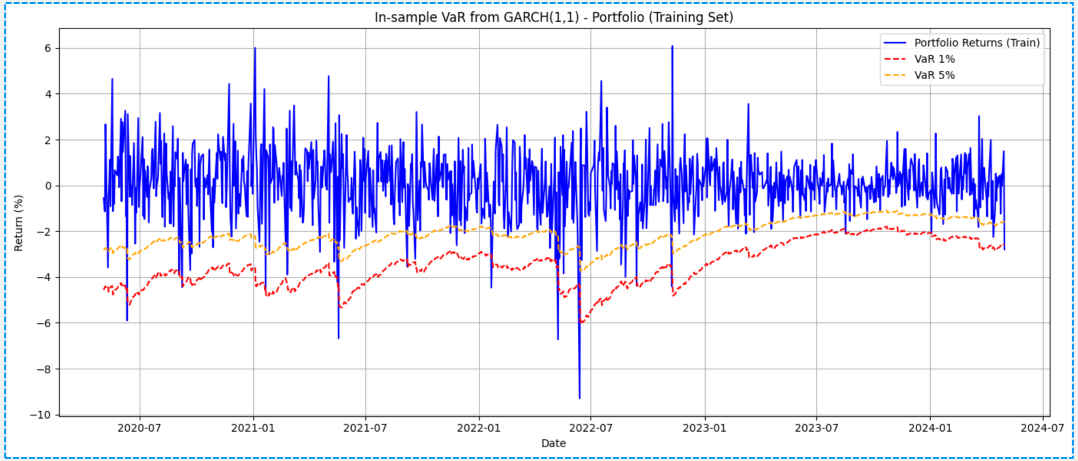


On the train set we are estimating the VaR using the GARCH(1,1) fitted before

$$\sigma_t^2 = 0.0071 + 0.0310 \cdot u_{t-1}^2 + 0.9648 \cdot \sigma_{t-1}^2 \quad \text{where } u_t \sim t(6.55)$$

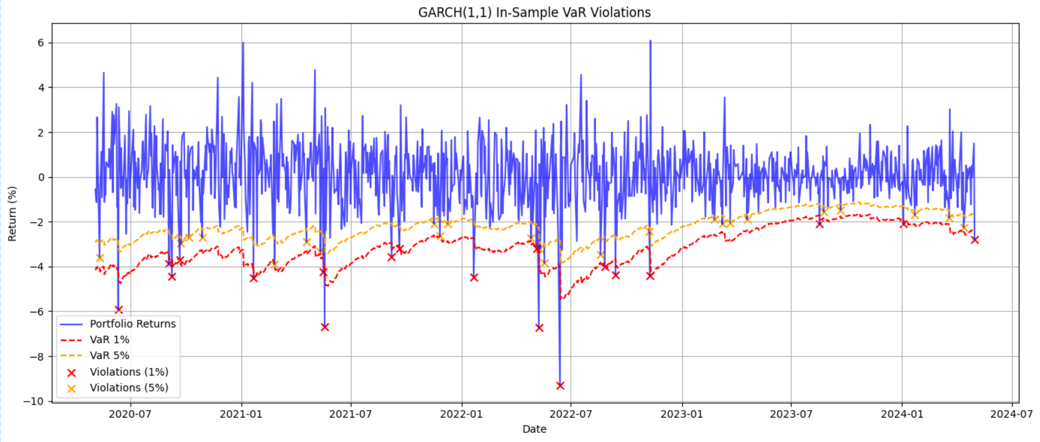
$\omega = 0.0071$ $\alpha = 0.0310$ $\beta = 0.9648$ $\nu = 6.55$ degrees of freedom.

VaR in-sample - GARCH(1,1) model



The plot above compares the portfolio returns (in blue) with the estimated in-sample Value-at-Risk (VaR) at the 1% and 5% levels.
By visualizing the VaR alongside the portfolio returns, we can assess the accuracy of the model in capturing potential tail risks.

VaR in-sample violations - GARCH(1,1) model



We expect approximately:

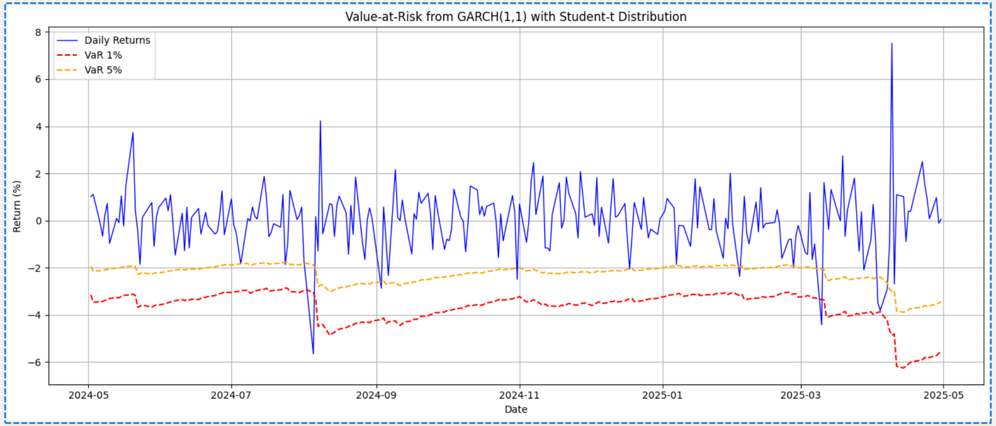
1% of violations for the 1% VaR,
5% of violations for the 5% VaR.

Our results:

VaR 1% Violations: 15 out of 1005 observations (1.49%)
VaR 5% Violations: 50 out of 1005 observations (4.98%)

A violation rate slightly above 1% for the 1% VaR is acceptable and expected in practice. The 1% VaR threshold is very narrow and targets extreme tail events.

VaR test set forecasting - GARCH(1,1) model



The GARCH(1,1) model is **fitted once** on the in-sample period and kept fixed.
For each day in the out-of-sample period:

- A **rolling window** (1005 obs*) of past returns is used. → perfect balance between noise avoidance and loss of info
- **New residuals** are computed using the fixed mean μ.
- The GARCH recursion is applied across the entire window to recalculate the conditional variance at the end of the window. This variance is used to *forecast 1-day ahead VaR*, without re-estimating model parameters.

**Decision taken in line with Basel III recommendations, which suggest using at least one year of historical data for market risk estimation. Other studies suggest a rolling window of 1000 elements (source)*

Thanks!

Do you have any questions?

g.fantato@student.uw.edu.pl | f.mennella@student.uw.edu.pl
d.castaldo@student.uw.edu.pl

GitHub Project link