Mid term Exam for Financial Econometrics with Python

PRAT Paul; GAVINI Charles; FOURNIER Justin; BLANC Mathieu

November 14, 2024

Contents

1 Introduction

This document provides a comprehensive presentation of our results, including all relevant tables, figures, and calculations. The report is structured into distinct parts, beginning with the importation of essential Python libraries. We then initialize variables to organize the data into different categories (e.g., daily, monthly, returns, log returns), allowing for clear analysis and comparison across various data types and intervals.

2 Preliminary

2.1 AMAZON

The selected stock for this analysis is Amazon due to its significant relevance in current global markets, its impressive growth over time and its position as a major industry leader. The ticker from yahoo finance is "***AMZN*** " on the Nasdaq stock exchange AMAZON on Yahoo Finance First, importing the Amazon stock with yfinance, then display the pandas table. We will import 25 years, 8 months and 25 days of data (from 1999-01-21 to 2024-10-16).

2.2 Data Table

The data printed here is the preview of the Amazon stock extraction from yahoo finance:

	Open	High	Low	Close	Adj Close	Volume
Date						
1999-01-21	2.612500	2.759375	2.314063	2.650000	2.650000	940964000
1999-01-22	2.487500	3.146875	2.468750	3.075000	3.075000	875316000
1999-01-25	3.037500	3.084375	2.750000	2.809375	2.809375	546476000
1999-01-26	2.815625	3.031250	2.765625	2.877344	2.877344	490696000
1999-01-27	3.353125	3.493750	3.000000	3.140625	3.140625	700452000

Table 1: Preview of Amazon Stock Data (5 first datas) from "AMZN" in Yahoo Finance

2.3 Checking the 25 Years range condition

We need to verify that the data displays accurately over the 25 years range. Fortunately, the extracted Amazon data has been available since January 1999. To ensure the data's continuity and completeness, we will implement a Python script that identifies and counts any gaps within the dataset. By visualizing the dates of these gaps, we can easily detect any significant interruptions that could potentially impact our data analysis

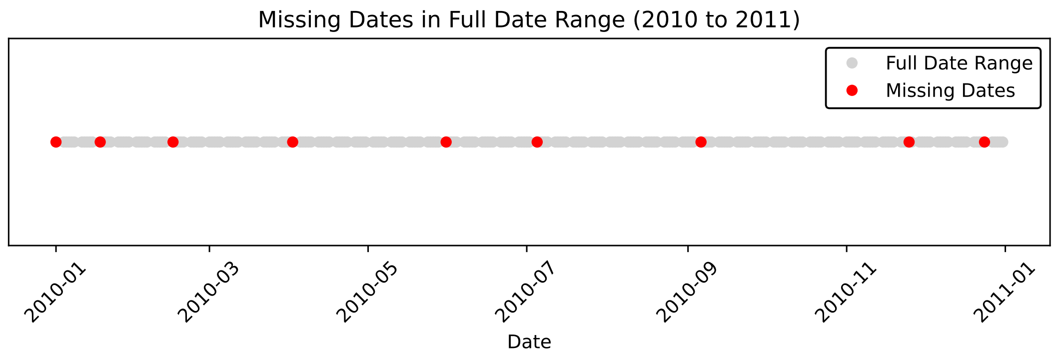

Figure 1: Missing Dates in a partial date range (**01-01-2010** to **01-01-2011**)

We identified a total of 238 isolated days of data gaps per year across the 25-years range (6476 values). Therefore, the data remains reliable for our stylized facts analysis. The missing data points in our dataset are randomly distributed and account for 3.7% of the total data. According to scientific studies on data reliability for volatility testing, a dataset with up to 10 [2]

3 First Results

3.1 Prices Evolutions

With the accuracy and the reliability of our dataset confirmed, we begin by plotting the evolution of prices over 4 different periods : Daily, Weekly, Monthly and Yearly prices.

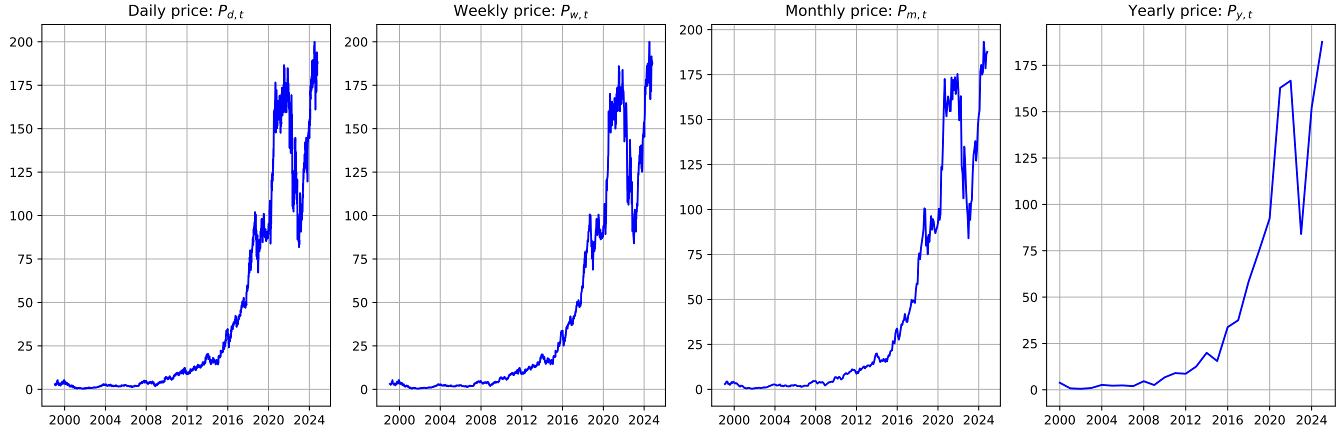

Figure 2: Prices over time P_t by frequency daily, weekly, monthly and annual the AMZN stock. Sample: **01-21-1999** to **10-16-2024**.

3.2 Calculating Returns

Using the processed data, we can now output graphs for several key metrics: daily prices, daily log prices, daily simple returns, and daily log returns. Plotting these metrics will allow us to observe daily price movements, the transformation of prices into *log* form for trend analysis, as well as daily returns and their logarithmic equivalents.

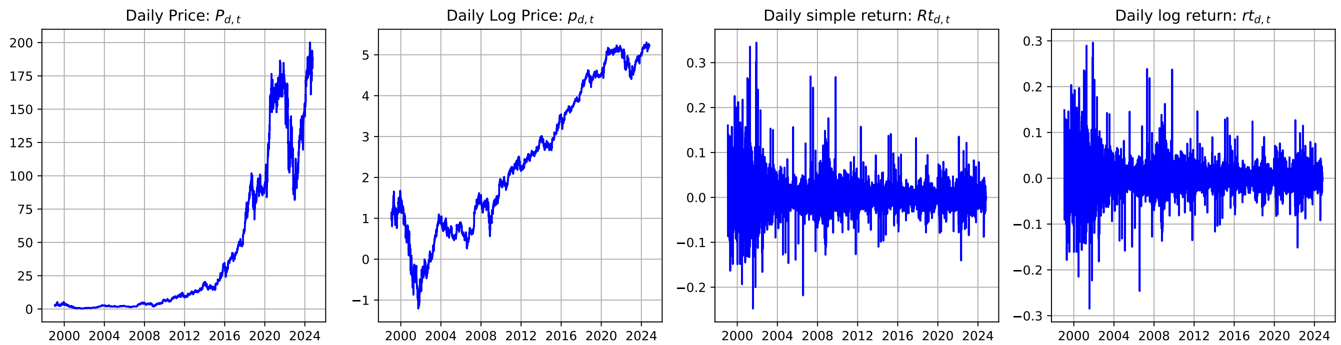

Figure 3: Prices P_t, returns R_t and log returns r_t of the AMZN stock. Sample: **01-21-1999** to **10-16-2024**.

4 Amazon and the 8 Stylized Facts

4.0.1 Summary statistics

	daily	weekly	monthly	annual
Mean	0.06351	0.31014	1.32164	22.85692
St.Deviation	3.24131	6.76830	13.06275	45.28052
Diameter.C.I.Mean	0.07896	0.36213	1.45887	18.11598
Skewness	0.41992	0.05008	-0.45895	-0.15137
Kurtosis	11.15158	7.60655	2.59462	-0.64793
Excess.Kurtosis	8.15158	4.60655	-0.40538	-3.64793
Min	-28.45678	-38.51804	-53.02674	-68.54809
Quant5	-4.61051	-9.74288	-20.16713	-55.72274
Quant25	-1.25994	-2.64062	-4.98163	-7.23999
Median	0.04108	0.30519	2.09626	23.07665
Quant75	1.39659	3.40897	8.45973	55.96192
Quant95	4.47118	10.67416	20.90661	94.77653
Max	29.61811	56.11507	48.35221	102.44636
Jarque.Bera.stat	33735.75720	3235.87866	97.20696	0.51147
Jarque.Bera.pvalue.X100	0.00000	0.00000	0.00000	77.43487
Lillie.test.stat	0.10194	0.09591	0.08194	0.06494
Lillie.test.pvalue.X100	0.10000	0.10000	0.10000	99.00000
N.obs	6474.00000	1342.00000	308.00000	24.00000

Table 2: Summary statistics for the AMZN stock. Sample: **01-21-1999** to **10-16-2024**.

4.1 Prices are non-stationary

The first feature that will highlight non-stationarity of the prices is the comparison of p_t vs p_{t-1}.

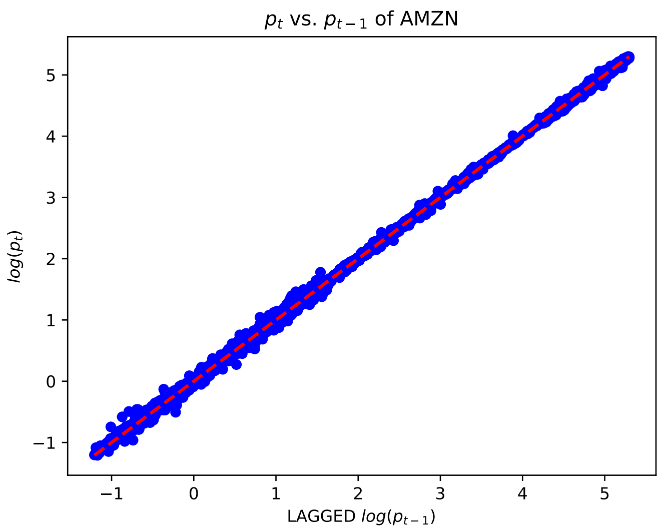

Figure 4: Comparison of $\log(p_t)$ vs $\log(p_{t-1})$ of the AMZN stock. Sample: **01-21-1999** to **10-16-2024**.

The graph in Figure 4 demonstrates this strong linear relationship, indicating that Amazon's prices at time t are highly dependent on those at $t-1$ and lack mean reversion, supporting the idea of non-stationarity.
Additionally, the empirical autocorrelation function (ACF) of Amazon's daily prices shows a slow decay, further suggesting non-stationarity, as shown in the next figure.

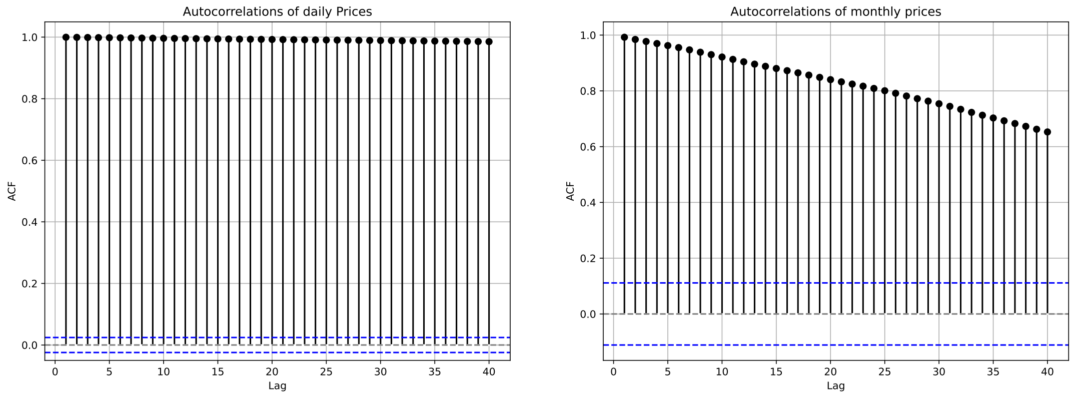

Figure 5: Autocorrelations of daily and monthly prices of the AMZN stock. Sample: **01-21-1999** to **10-16-2024**.

For the amzon daily and monthly prices time series, we expect to see large values of $\hat{\rho}_k$, i.e., near to 1, slowly decaying as k increases this is the **long memory property**. Also, the **slowly decaying ACF** is often a symptom of non stationarity.

4.2 Returns are stationary

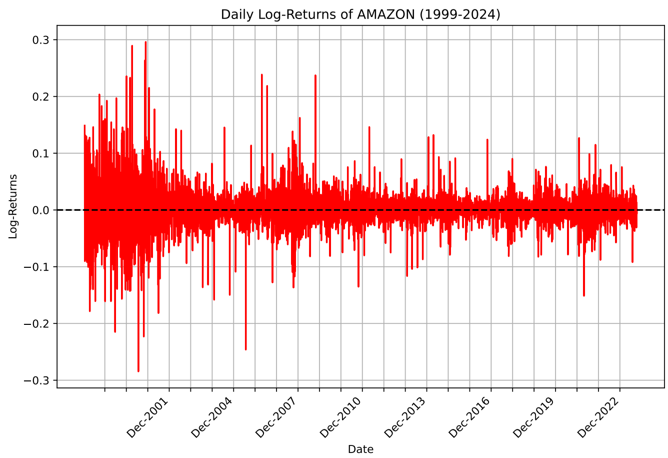

Figure 6: Daily Log-returns $r_t := p_t - p_{t-1}$ of the AMZN stock. Sample: **01-21-1999** to **10-16-2024**.

Log-returns are a common way to measure the percentage change in stock prices, and they help assess the stability or stationarity of the returns over time. In a stationary series, we would expect the properties, such as mean and variance, to remain constant over time. However, here we observe significant differences in volatility across the timeline.

In the early years (around 1999-2005), there is noticeably higher volatility in Amazon's log-returns, with frequent large spikes both upwards and downwards. This period corresponds to the tech boom and subsequent dot-com bubble burst, during which many tech stocks, including Amazon, experienced extreme price fluctuations. Additionally, as a relatively new and fast-growing company, Amazon's stock likely faced higher uncertainty and speculative trading, contributing to greater volatility.

4.3 Asymmetry

	daily	weekly	monthly	annual
Skewness	0.41992	0.05008	-0.45895	-0.15137
Kurtosis	11.15158	7.60655	2.59462	-0.64793

Table 3: **Skewness and kurtosis of daily, weekly, monthly and annual** *log* returns of the AMZN stock. Sample: **01-21-1999** to **10-16-2024**.

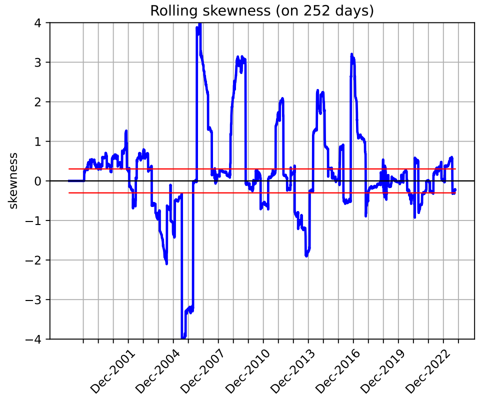

Figure 7: **Rolling skewness** of AMZN stock. Sample: **01-21-1999** to **10-16-2024**. The red bands corresponds to the limit of acceptance, the blue line correspond to the rolling skewness with T=252

For this case, Table 3 hilights that for daily returns the AMZN stock skewness is positive. This does not confirms stylized fact 3, this case is not really common but it means that the mean return is higher than the median of the sample [1]. Then, Amazon investors tend to have steeper high turns than downturns and that investors of the AMZN stock react more positively to good news than they can react badly for negative news. If we take a look at the Rolling Skewness on simple returns Figure 7, we clearly see that the skewness (for a 252 days interval) varies a lot depending the position of intervals and has already been very negative (Dec 2004) but is generally positive.

4.4 Heavy tails

As showcased in the Table 3, there is a large excess kurtosis

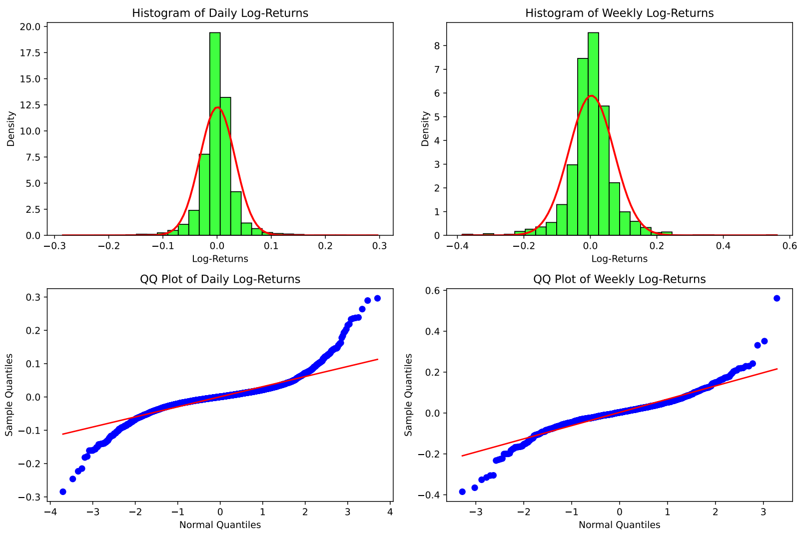

Figure 8: Log returns $r_t := p_t - p_{t-1}$: **histograms of daily, monthly** "adjusted closing" of AMAZON. Sample: **01-21-1999** to **10-16-2024**. QQ plot against quantiles of normal distribution with same mean and variance as the empirical distribution of returns.

Here, the QQ-Plots explicit clearly how our sample distinguishes from the normal distribution. The QQ-plot provide graphical evidence that the tails of the daily returns distributions are heavier than the tails of the normal distribution as: The points on the left of the graph which represent the lower quantiles (i.e. the points in the left tail of the empirical distribution) are below the blue line. The lower quantiles of the empirical distribution are much smaller than what you should expect from a Normal random variable with the same empirical mean and standard deviation of the sample the left tail of the empirical distribution is heavier than one of a Normal Distribution. similar conclusions for the right tail.

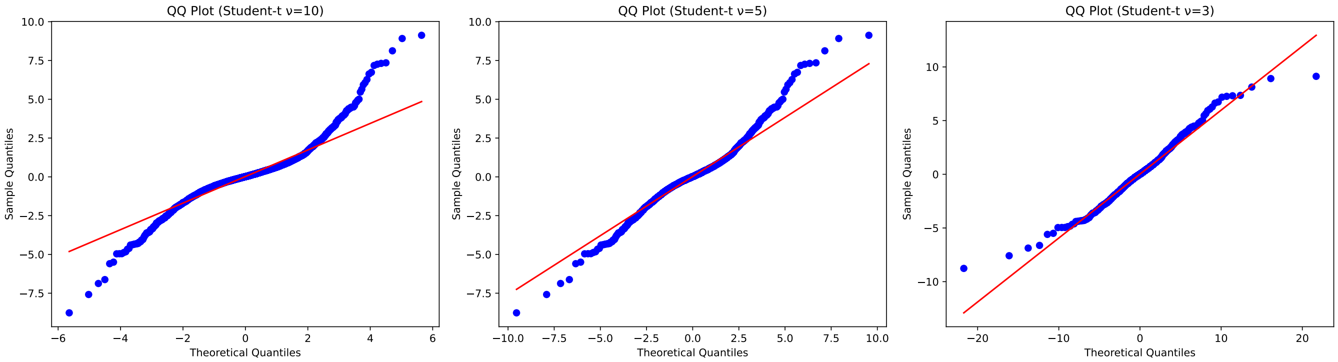

Figure 9: Log returns $r_t := p_t - p_{t-1}$: **daily** "adjusted closing" of AMZN stock. Sample: **01-21-1999** to **10-16-2024**. **QQ plot of Sample standardized quantiles** (0 mean and unit variance) **of daily log-returns against quantiles of standardized** (0 mean and unit variance) **Student-t distributions** with $\nu = 10$, 5, and 3 degrees of freedom.

4.5 Gaussianity

4.5.1 High frequency non-Gaussianity

The aggregate gaussianity, states that lower frequency returns (monthly) tend to be Gaussian (symmetric about the mean) even if higher frequency returns (daily) are not. To test this stylized fact we perform a Jarge-Bera test. The result is in Table 2.

The 3$^{\text{rd}}$ central moment is defined as $\mu_3 := E((X - m_1)^3)$. The skewness of r_t is defined as:

$$\text{Skew}(r_t) := E\left[\left(\frac{X - m_1}{\sigma}\right)^3\right] = \frac{\mu_3}{\sigma^3} = \frac{\mu_3}{\mu_2^{3/2}}.$$

As the result in Table 2 the skewness is positive for daily and monthly data, $\text{Skew}(r_t) > 0$, large realizations of X are more often larger than the mean μ. Skewness is thus used as a measure of asymmetry of the distribution $f_X(x)$. Therefore:

Skew$(r_t) > 0$, so the distribution is said to be **right skewed**.
Skew$(r_t) > 0$, then $\mu >$ median.

4.5.2 Aggregational Gaussianity

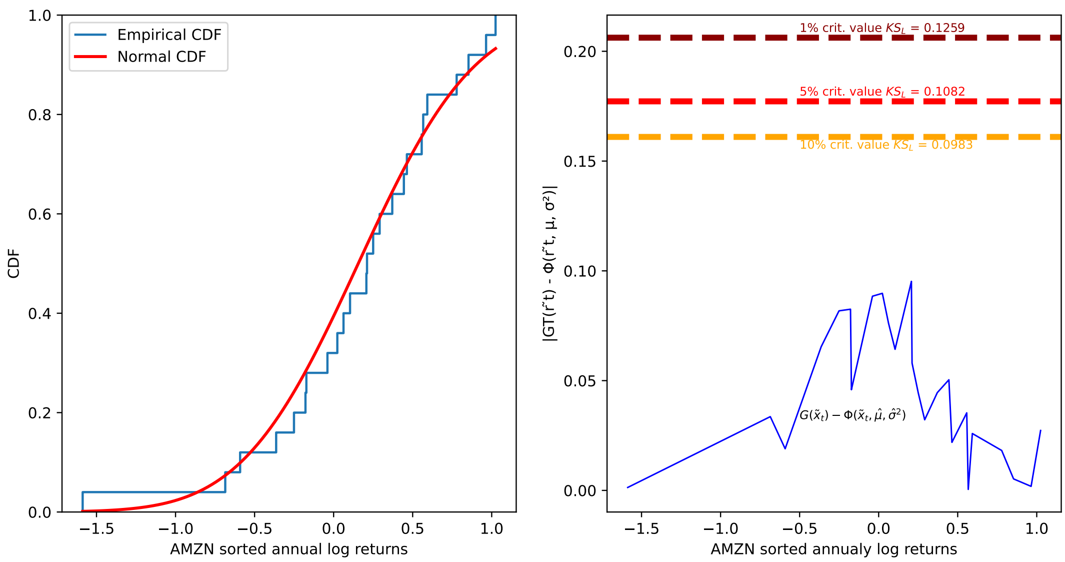

Figure 10: Log returns $r_t := p_t - p_{t-1}$: **annual** "adjusted closing" of AMZN. Sample: **01-21-1999** to **10-16-2024**. **Left panel**: empirical and Normal cdf's for the standardized annual returns of AMZN. **Right panel**: values of $\left|G_T(\tilde{r}_t) - \Phi(\tilde{r}_t, \hat{\mu}, \hat{\sigma}^2)\right|$ (blue line) and critical values for the Lilliefors test for the three significance levels 10%, 5% and 1%.

The blue line is under the critical values lines, So the test is respected and so for the Gaussianity.

4.6 Returns are not autocorrelated

Stylised fact 6 posits that returns are not autocorrelated. Autocorrelation in a weakly stationary process measures the correlation between values of the process at different time points. To assess the significance of autocorrelations, we apply the **Box-Pierce (BP)** test or the **Ljung-Box (LB)** test, where the null hypothesis indicates that all autocorrelations are equal to 0, compared with the alternative analysis where it differs from 0.

	lag	acf	acf diam.	acf test	B-P stat	B-P pval	L-B stat	L-B pval	crit
0	1	0.053	0.112	0.930	0.865	0.352	0.874	0.350	3.841
4	5	0.021	0.112	0.376	7.654	0.176	7.796	0.168	11.070
14	15	-0.049	0.112	-0.865	22.975	0.085	23.673	0.071	24.996
24	25	-0.048	0.112	-0.846	36.101	0.070	37.859	0.048	37.652

Table 4: Ljung-Box and Box-Pierce daily

In Table 4, we observe that for each lag (1, 5, 15 and 25), the p-values for both the **Ljung-Box (L-B pval)** and **Box-Pierce** ($B-Ppval$) tests exceed the 0.048 **threshold**. This indicates that there is insufficient statistical evidence to reject the null hypothesis at each lag. Consequently, this suggests that the returns are not significantly autocorrelated across these time lags.

4.7 Volatility clustering and long range dependence of squared returns

Volatility clustering is a phenomenon where periods of high market volatility are often followed by high volatility, and vice versa. To capture and analyze this phenomenon, financial models such as ARCH (Autoregressive Conditional Heteroskedasticity) and GARCH are commonly used. We can easily perceive it on the graph below, (from december 2001 to december 2004) phase of low volatility.

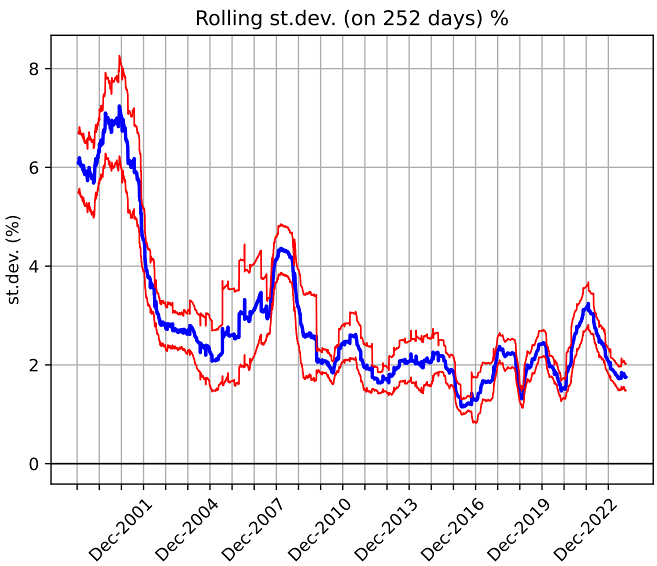

Figure 11: Rolling standard deviation from the “adjusted closing” of AMZN. Sample: **01-21-1999** to **10-16-2024**.

This persistence in the autocorrelation of squared returns reflects volatility clustering. High volatility often persists over time before settling into a lower volatility regime; this is how time dependance is reflected.

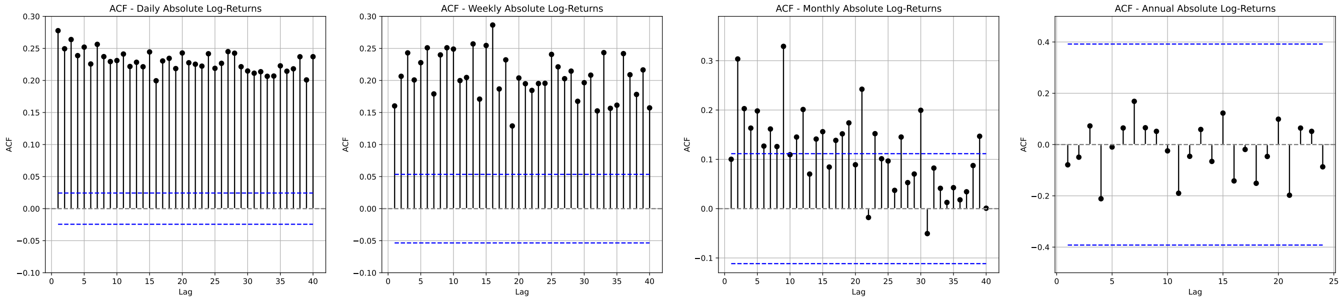

Figure 12: Autocorrelations of the daily, weekly and monthly absolute log-returns r_t from the "adjusted closing" of AMZN. Sample: **01-21-1999** to **10-16-2024**. The blue dotted bands represents the confidence intervals (**Barlett intervals**), $\frac{1}{\sqrt{T}}$ where T is the number of samples.

We observe that the autocorrelation is continuous, as indicated by the trendline, which aligns with the previous graph. Additionally, it becomes apparent that as the time interval changes (from daily to weekly, monthly, and annually) the autocorrelation becomes more pronounced between intervals. This aligns with the volatility clustering phenomenon discussed earlier. This effect occurs because ARCH and GARCH models are sensitive to sampling frequency, with their impact being more noticeable at shorter frequencies (daily, weekly) than at longer ones (monthly, annually).

4.8 Leverage effect

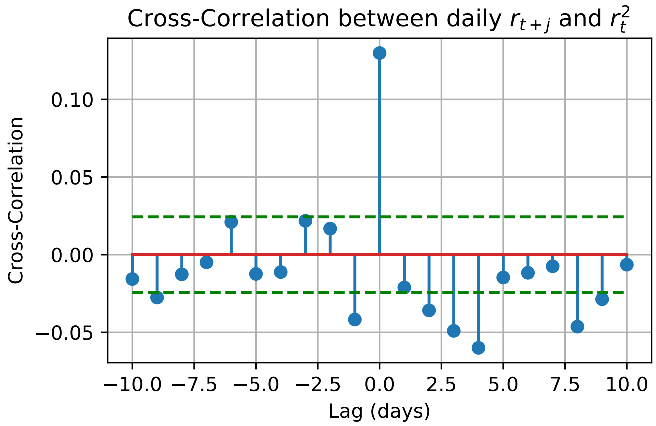

Figure 13: Rolling standard deviation from the "adjusted closing" of AMZN. Sample: **01-21-1999** to **10-16-2024**.

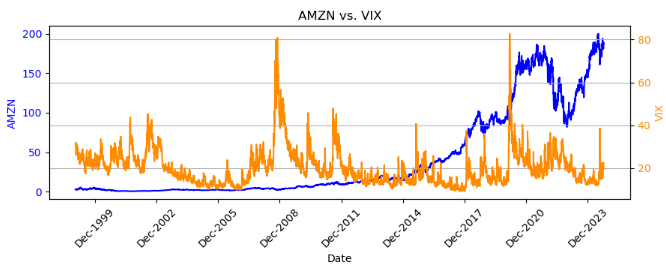

Figure 14: Rolling standard deviation from the "adjusted closing" of AMZN. Sample: **01-21-1999** to **10-16-2024**.

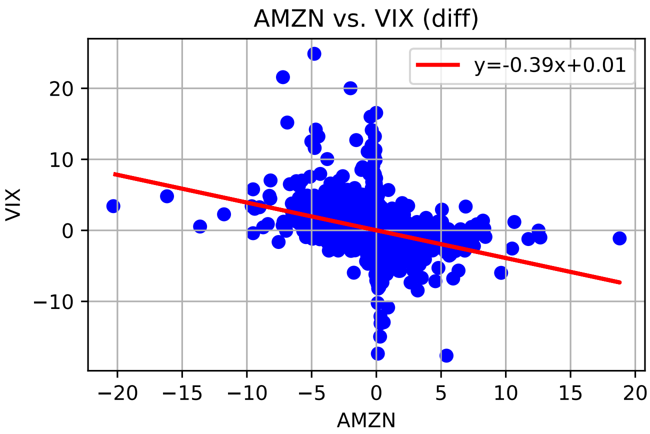

Figure 15: Rolling standard deviation from the "adjusted closing" of AMZN. Sample: **01-21-1999** to **10-16-2024**.

A Appendix: Python Code

Below is the Python code used in this analysis.

```
# Python code example
import numpy as np
import pandas as pd

def analyze_data(data):
    mean = np.mean(data)
    std_dev = np.std(data)
    return mean, std_dev

data = [1, 2, 3, 4, 5]
mean, std_dev = analyze_data(data)
print(f"Mean: {mean}, Standard Deviation: {std_dev}")
```

Listing 1: Python Code for Analysis

B To go further, CAPM pricing model

References

[1] Rui Albuquerque. Skewness in stock returns: Reconciling the evidence on firm versus aggregate returns. *The Review of Financial Studies*, 25(5):1630–1673, May 2012. Published: 09 January 2012.

[2] Giovanni Pumi et al. Estimation of long-range dependent models with missing data: to impute or not to impute? *arXiv preprint*, 2023.