UNIVERSITÀ DEGLI STUDI DI PADOVA

Università degli Studi di Padova

Master Degree Course in Computational Finance

Regression and Time Series Models

Group Work 1 - Regression with CAPM Model

Created by:

Martina Arrighini, 2149799, martina.arrighini@studenti.unipd.it

Luigi Babiski Arruda, 2163341, luigi.babiskiarruda@studenti.unipd.it

Giorgio Cottini, 2140740, giorgio.cottini@studenti.unipd.it

Enrico Paciaroni, 2139678, enrico.paciaroni@studenti.unipd.it

Commissioned by:

Prof. Massimiliano Caporin

A.Y. 2024/2025

Contents

Overview

This report analyzes the Healthcare sector within the Euro Stoxx 600 index, using data from October 2000 to October 2024 to evaluate the applicability of the Capital Asset Pricing Model (CAPM). The Healthcare sector, known for its inelastic demand, high entry barriers, and resilience, provides a compelling context for assessing risk-return dynamics.

Using Ordinary Least Squares (OLS) regression, we estimate CAPM parameters, including alpha, beta, and R-squared, for both individual equities and an equally weighted portfolio. Diagnostic tests, such as the White and Breusch-Godfrey tests, ensure the robustness of the model. Additionally, a multifactor model based on the Fama-French framework is introduced to compare its explanatory power with that of the CAPM.

Structural breaks in the data are analyzed using the Chow test and a rolling window approach, providing insights into parameter stability over time. This collaborative effort aims to deliver a comprehensive evaluation of the CAPM and multifactor models, offering valuable insights into the financial dynamics of the Healthcare sector.

Assignment 1

Data Choice and Download

In this assignment, we gathered and prepared the necessary data for analyzing the Healthcare sector within the Euro Stoxx 600 index. The data includes total return indices for individual stocks, risk-free interest rates, and the market index for the healthcare sector.

1.1 Data Collection

- **Total Return Indices:** We obtained the total return indices for ten companies within the Healthcare sector. These companies were selected based on the availability of complete data from 2000 to 2024.
- **Market Index:** The Euro Stoxx 600 Health Care index was used as a benchmark for the overall market performance within the sector.
- **Risk-Free Rates:** Monthly risk-free interest rates were derived from annualized Federal Reserve data to compute excess returns.

1.2 Data Preparation

- **Return Calculation:** Monthly returns were calculated using logarithmic differences in prices. This method ensures that the data is approximately normally distributed and facilitates economic interpretation.

- **Excess Returns:** Stock and market returns were adjusted by subtracting the risk-free rate, consistent with the CAPM framework.
- **Data Quality Checks:** The datasets were thoroughly examined for missing or duplicate values to ensure the robustness of subsequent analyses.

1.3 Preliminary Visualization

Scatter plots were created to examine the relationship between stock excess returns and market excess returns. These plots showed a clear linear relationship for most equities, supporting the applicability of the CAPM. However, some observations indicated potential outliers, which will be analyzed further in later steps.

Assignment 2

Relationship Between Stock and Market Returns

2.1 Linear Relationships and the CAPM Model

The scatter plots demonstrate a clear linear relationship between the excess returns of several stocks and the sector market returns, particularly for Johnson & Johnson, Pfizer, and Medtronic. This alignment suggests a strong correlation, with the CAPM model effectively describing the variability in their returns. The high R^2 values and significant β coefficients observed for these stocks reinforce the reliability of the CAPM model in explaining their behavior. These findings validate the assumption that systematic risk plays a crucial role in driving stock returns within the Healthcare sector. The consistency of these linear relationships highlights the applicability of the CAPM framework for stocks with stable market exposure. The regression lines for these companies align closely with the observed data points, indicating that their returns are predominantly influenced by market movements. This makes the CAPM model a reliable tool for estimating expected returns and assessing risk for these stocks. However, it is worth noting that while the CAPM model effectively explains returns for most stocks, it operates under the assumption of linearity and market efficiency. Deviations from these assumptions could impact its reliability, particularly for stocks influenced by unique, company-specific factors. As a result, further exploration of non-linear models or additional risk factors might be warranted for certain cases.

2.2 Weaker Correlations and Unique Factors

In contrast, stocks like Boston Scientific and Cigna exhibit weaker correlations with the sector market, as seen in their more dispersed scatter plot data points. This suggests that their returns are influenced by factors beyond those captured by the CAPM model, such as company-specific events, market segmentation, or operational inefficiencies. These deviations reduce the predictive power of the CAPM for these stocks. The weaker relationship for these companies could reflect the presence of idiosyncratic risk or external shocks that are not accounted for by the model. For example, Boston Scientific may be impacted by fluctuations in demand for its medical devices, while Cigna's returns could be shaped by changes in the regulatory landscape or shifts in the healthcare insurance market. These factors highlight the limitations of a purely market-based approach to risk estimation. As a result, while the CAPM model provides a foundation for understanding market-driven returns, it may need to be supplemented with additional analyses. Incorporating company-specific metrics or exploring multi-factor models could provide a more comprehensive understanding of the risks and returns associated with these stocks.

2.3 Outliers and Non-Linear Trends

The presence of outliers in the scatter plots of Figure 2.1 of stocks such as Revvity and Labcorp Holdings points to the influence of extraordinary events or structural breaks. These outliers may be linked to regulatory changes, breakthroughs in research and development, or unexpected market events, which introduce noise into the CAPM framework. Such deviations challenge the assumption of stable relationships between excess returns and market movements. Outliers can also indicate the presence of leverage or operational risks specific to certain companies. For example, regulatory approvals or setbacks for new products may cause sudden spikes or drops in stock performance that deviate from the expected trendline. These unique events underscore the importance of monitoring industry-specific developments when applying the CAPM model. To address these challenges, future analyses could incorporate robust techniques to manage outliers, such as data winsorization or robust regression methods. Additionally, exploring time-varying β coefficients or multi-factor models could provide a more nuanced understanding of how unique events impact stock returns.

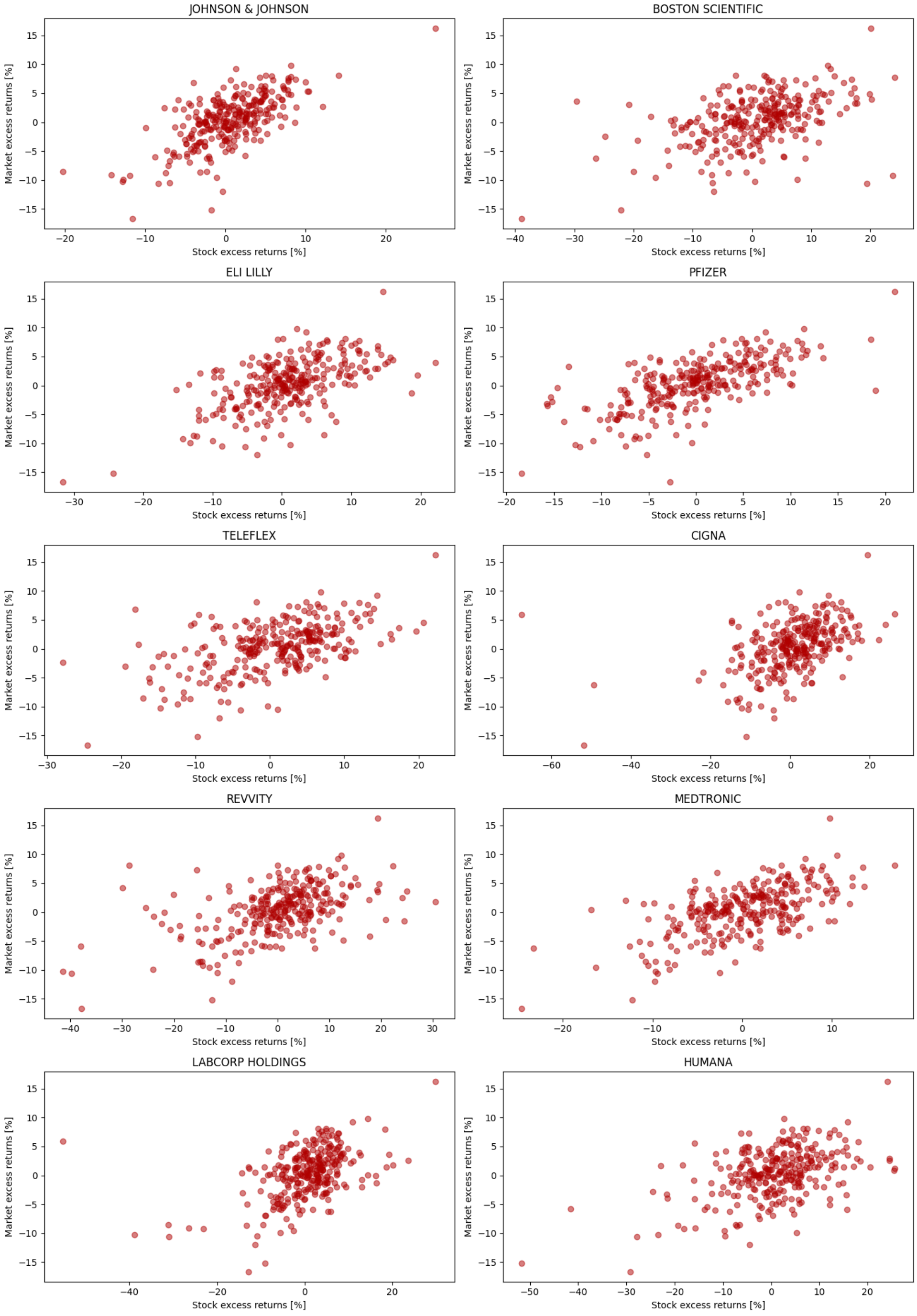

Fig. 2.1: Scatterplot of equities' returns against excess market returns, with linear regression.

Assignment 3

CAPM Regression

3.1 Regression Results for Individual Stocks

The CAPM regression applied to individual stocks highlights significant findings regarding α, β and R^2 values (see Table 3.1 for reference). Stocks like Johnson & Johnson and Pfizer exhibit high β coefficients, indicating a strong sensitivity to market movements, with market returns explaining a substantial portion of their variability. These results confirm the applicability of the CAPM model to these stocks and its ability to estimate systematic risk effectively. However, not all stocks perform equally well under the CAPM framework. Stocks like Cigna and Labcorp Holdings show relatively low R^2 values, indicating that a smaller proportion of their return variability is explained by market factors. This suggests the presence of idiosyncratic risks or company-specific drivers that are not captured by the model. These discrepancies highlight the limitations of CAPM in explaining returns for stocks with unique business models or exposure to niche markets. For stocks with lower R^2 values, alternative approaches, such as incorporating additional explanatory variables (e.g., size, value, or momentum factors), may enhance model reliability. These findings suggest that while the CAPM model provides a robust foundation for analyzing systematic risk, its assumptions may need to be adjusted for certain stocks.

Company	α	β	R^2	**Err.** α	**Err.** β	$\Delta\beta$ **[0.025]**	$\Delta\beta$ **[0.975]**
Johnson & Johnson	0.2451	0.7474	0.455	0.212	0.048	0.652	0.843
Boston Scientific	0.3464	0.9036	0.207	0.458	0.105	0.698	1.109
Eli Lilly	0.5655	0.9314	0.342	0.335	0.076	0.781	1.082
Pfizer	-0.3246	0.9595	0.451	0.274	0.063	0.836	1.083
Teleflex	0.3055	0.9621	0.297	0.383	0.088	0.790	1.134
Cigna	0.2511	1.0512	0.212	0.525	0.120	0.815	1.287
Revvity	-0.2457	1.2225	0.265	0.527	0.120	0.985	1.460
Medtronic	-0.1108	0.8613	0.384	0.282	0.065	0.734	0.988
LabCorp Holdings	0.2675	0.9163	0.233	0.430	0.098	0.723	1.110
Humana	0.5939	1.0700	0.231	0.506	0.116	0.843	1.298
Portfolio	**0.2234**	**0.8677**	**0.772**	**0.122**	**0.028**	**0.813**	**0.923**

Table 3.1: Regression results for each company, detailing Alpha, Beta, R-squared, and confidence intervals.

3.2 Performance of the Weighted Portfolio

The weighted portfolio demonstrates a strong alignment with the CAPM model, as evidenced by its significant and consistent β coefficient and high R^2 values. The compact and regular distribution of data points around the regression line highlights the effectiveness of diversification in reducing idiosyncratic risks and enhancing the model's explanatory power. This finding underscores the stability of portfolio returns when individual stock-specific risks are averaged out. Compared to individual stocks, the portfolio benefits from a smoother relationship with market returns, as diversification minimizes the noise caused by unique company events. This results in more reliable estimates of β and a higher goodness of fit (R^2), confirming the portfolio's sensitivity to systematic risk. These characteristics make the CAPM model particularly suitable for analyzing aggregate portfolio behavior. The portfolio's performance under the CAPM model reinforces the importance of diversification as a strategy for mitigating risk and achieving more predictable returns. By averaging out the effects of outliers and company-specific shocks, the portfolio provides a clearer view of systematic risk and market sensitivity.

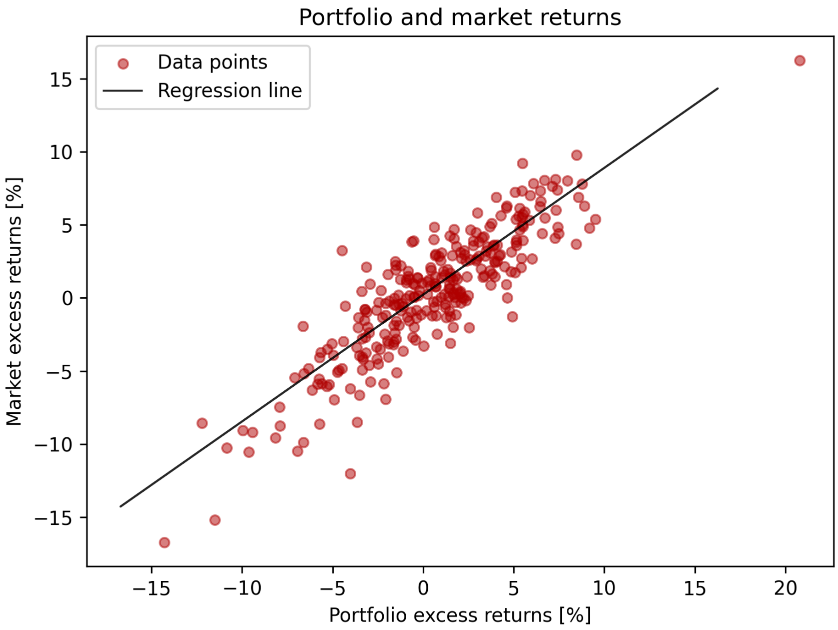

Fig. 3.1: Scatterplot of portfolio's returns against excess market returns, with linear regression.

3.3 Impact of Diversification

Diversification proves to be a key factor in stabilizing the parameter estimates of the CAPM model. While individual stocks may exhibit high variability in β and α due to idiosyncratic risks, the equally weighted portfolio achieves greater consistency. This is particularly evident in its reduced residual variance and higher R^2 values, which indicate a stronger fit to the CAPM framework. The ability of diversification to average out company-specific risks highlights its role in improving the reliability of financial models. For investors, this suggests that constructing diversified portfolios can lead to more predictable risk and return profiles, even when individual stocks deviate from expected patterns. Diversification also reduces the influence of outliers, providing a clearer picture of systematic risk. These observations underscore the practical applications of the CAPM model for portfolio analysis, where diversification enhances its explanatory power. However, the model's limitations in capturing unique factors for individual stocks emphasize the need for supplemental analyses when evaluating less diversified assets.

Assignment 4

Diagnostic Tests

4.1 Significance of β and R^2 Values

The results of the CAPM regression, displayed in Table 4.1, confirm the model's ability to explain the returns of many stocks. Stocks like Johnson & Johnson (β=0.752, R^2=0.458) and Pfizer (β=0.953, R^2=0.449) exhibit significant β coefficients and high R^2 values, indicating strong sensitivity to market movements. These results validate the CAPM model's assumption that systematic risk is the primary driver of returns for these stocks. However, stocks with lower R^2 values, such as Cigna and Labcorp Holdings, suggest the presence of significant idiosyncratic risks or external factors that are not captured by the CAPM framework. These findings emphasize the model's limitations when applied to companies influenced by unique events or niche market dynamics. To improve the model's accuracy for such cases, additional factors, such as industry-specific risks or macroeconomic variables, could be incorporated. This would help account for variability that the CAPM model alone cannot explain, particularly for stocks with lower market correlations.

4.2 Heteroscedasticity Issues

The White Test identifies heteroscedasticity in stocks such as Johnson & Johnson and Humana, indicating non-constant residual variance. This violation of CAPM assumptions suggests that the reliability of the regression results may be compromised, as standard

Equity	β	β p-value	White p-value	BG p-value	R^2	HAC p-value
Johnson & Johnson	0.752	0	7.19E-10	0.178	0.458	0.001
Boston Scientific	0.91	0	2.17E-05	0.846	0.21	0.003
Eli Lilly	0.942	0	0.013	0.085	0.346	0.009
Pfizer	0.953	0	0.209	0.094	0.449	0.008
Teleflex	0.968	0	0.921	0.762	0.3	0.010
Cigna	1.056	0	0.119	0.988	0.214	0.015
Revvity	1.218	0	0.130	0.999	0.265	0.020
Medtronic	0.859	0	0.477	0.192	0.384	0.012
Labcorp Holdings	0.921	0	0.207	8.85E-07	0.236	0.005
Humana	1.081	0	2.87E-05	0.989	0.235	0.014

Table 4.1: Regression results for each equity, including Beta, p-values, and various test results.

errors may be underestimated or overestimated. To address this issue, robust standard errors (HAC) were applied, which adjust for heteroscedasticity and provide more reliable parameter estimates. These adjustments ensure that the significance levels of β coefficients remain valid despite violations of homoscedasticity. Future analyses could explore whether specific periods or events contribute to heteroscedasticity. Understanding these patterns may help refine the CAPM model's application to stocks affected by irregular variance in their residuals.

4.3 Autocorrelation in Residuals

The Breusch-Godfrey Test detects autocorrelation in stocks such as Labcorp Holdings, suggesting that the residuals are not independently distributed. This violates the CAPM assumption of no autocorrelation, which can lead to biased parameter estimates and reduced model reliability. To mitigate the impact of autocorrelation, adjustments such as incorporating lagged variables or alternative error correction methods can be applied. These techniques improve the robustness of the regression results and enhance the reliability of the CAPM model for stocks with autocorrelated returns. Despite these challenges, the CAPM model remains effective for analyzing diversified portfolios. The reduction of idiosyncratic risks and the alignment of residuals with model assumptions further support its applicability at an aggregate level. However, for individual stocks with significant deviations, additional diagnostic tools and model refinements are necessary to ensure accurate analysis.

Assignment 5

Chow Test

5.1 Structural Breaks and Chow Test

In this section, we employ the Chow test to assess whether a significant structural change occurs in the regression models at specific points in time. To ensure the robustness of our analysis, we first established a minimum data subset size for the unrestricted models, setting it to 10% of the total dataset, in order to maintain statistical validity while preserving sufficient data for meaningful comparisons. Subsequently, the Chow test was systematically applied to all linear regressions conducted on the selected equities, allowing us to identify potential structural breaks across the dataset, that is, significant changes in the regression parameters caused by external shocks, market-wide events, or company-specific factors. These structural breaks may reflect shifts in the relationship between excess returns and market behaviour, such as changes in systematic risk (β) or the presence of unexplained excess returns (α) due to macroeconomic conditions, regulatory adjustments or sector-specific developments. To identify these structural breaks, we analyzed the p-values obtained from the Chow test for each equity over time: a p-value below the threshold of 0.01 was interpreted as evidence of a structural break at that particular point in time, signaling that the CAPM parameters had changed significantly. In this analysis, only periods of at least two consecutive months with p-values below the threshold were considered as break dates, ensuring that the identified breaks reflect sustained shifts rather than noise or transient fluctuations.

The results of this analysis are displayed in Figure 5.1, highlighting that different

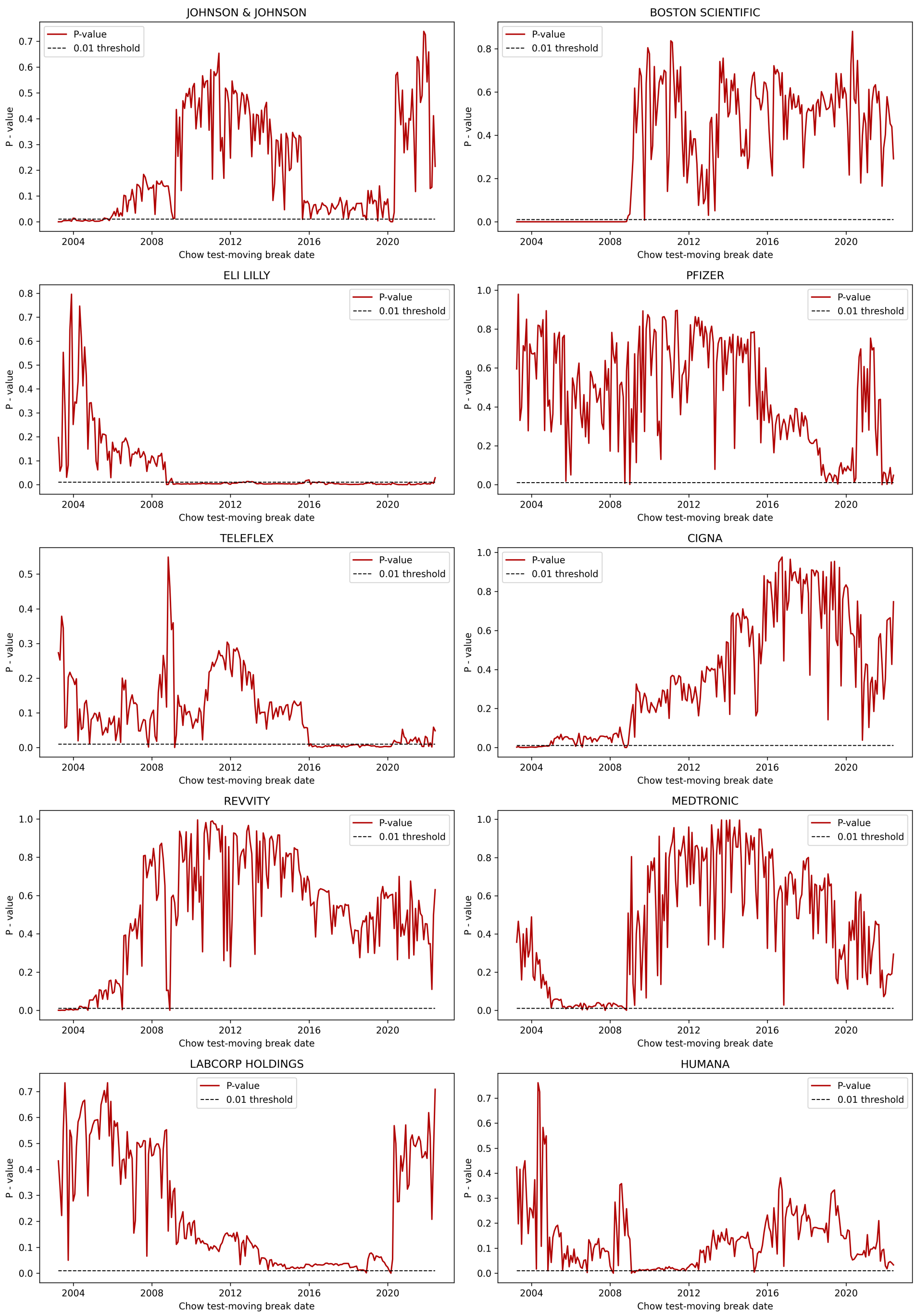

Fig. 5.1: Chow Test performed for all equities in search of structural breaks.

equities exhibit fundamentally distinct behaviors. Notably, certain equities, such as *Eli Lilly*, *Boston Scientific*, and *Teleflex*, show p-values consistently below the 0.01 threshold for extended periods, indicating prolonged structural breaks; this raises the question of whether the CAPM relationship for these equities in these periods is truly linear or whether a different modeling approach may be warranted.

5.2 Periods of Shared Structural Breaks

Despite these observations, no clear or consistent pattern of structural breaks emerges across all equities; to explore potential commonalities further, an additional histogram was generated (Figure 5.2), which illustrates the frequency and overlap of structural breaks shared among different equities.

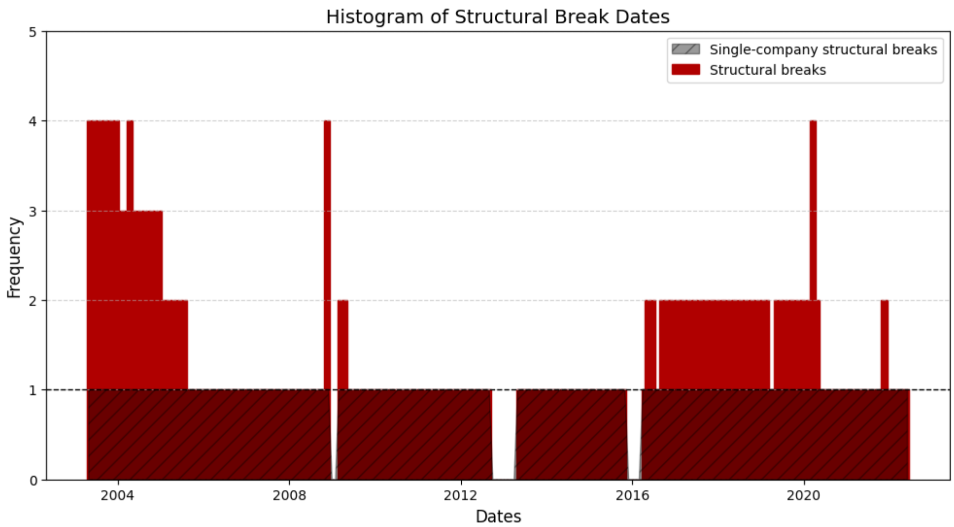

Fig. 5.2: Frequency of months identified as structural break points.

Figure 5.2 highlights two important aspects of the data: first, there is no single period that constitutes a break date for all equities; the maximum number of companies sharing a structural break at any given time is in fact 4. Nonetheless, the break dates identified as significant changes to the regression parameters align closely with expectations, in fact the data reveal break dates significantly relevant in three distinct periods of time:

- **2003**: Structural breaks are observed throughout most of the year, coinciding with the SARS epidemic. This period also aligns with economic and legislative developments, including the introduction of the Medicare Modernization Act, which reshaped healthcare policy and access in the United States.

- **October and November 2008**: These months mark the onset of the 2008/2009 financial crisis.

- **February and March 2020**: Structural breaks during this time correspond to the emergence of the COVID-19 pandemic and the widespread implementation of precautionary measures.

An important observation is that the structural breaks identified in 2003 extended over a prolonged period, spanning multiple consecutive months, whereas the impact of the COVID-19 pandemic appears to have been more concentrated and shorter in duration. The prolonged effect in 2003 could reflect the gradual adjustment of the healthcare market to the SARS epidemic and concurrent economic and legislative changes. In contrast, the shorter duration of structural breaks during COVID-19 may suggest that lessons learned from the SARS outbreak, including the implementation of precautionary measures and improved epidemic preparedness, helped mitigate the impact of subsequent epidemics on the healthcare market. Alternatively, the healthcare market may have simply grown so significantly in size and diversification over time that it developed a level of resilience that enables it to absorb and adapt to substantial disruptions in the economic landscape. This growth, coupled with advancements in technology, expanded infrastructure, and more robust financial mechanisms, might have contributed to the market's ability to withstand even major shocks, such as those caused by the 2008 crisis and the COVID-19 pandemic. This explanation also accounts for the observation that the majority of equities were not significantly impacted during these events: the increased resilience and diversification of the healthcare market may have allowed many companies to maintain stability despite the broader economic disruptions.

5.3 Chow Test on Portfolio

In order to further investigate the impact of significant events in the economic environment, the same Chow test was performed on the excess returns of the portfolio. This approach aggregates the behavior of individual equities into a single measure, allowing for the identification of systemic disruptions across the healthcare market.

The results, displayed in Figure 5.3, reveal that there is only one time period in which structural breaks in the portfolio's return dynamics are particularly significant: that is the beginning of 2009, likely reflecting the aftermath of the 2008 financial crisis.

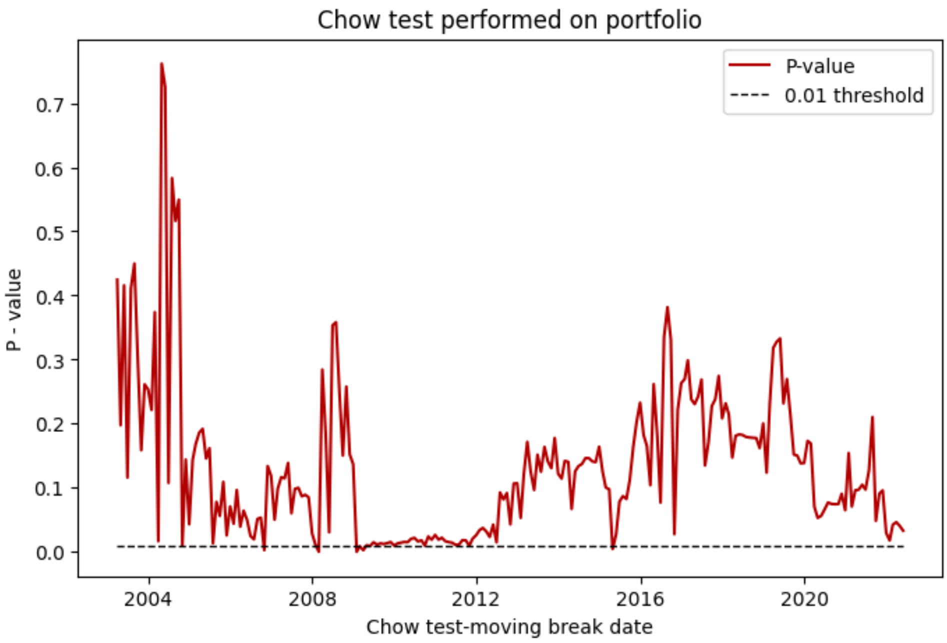

Fig. 5.3: Frequency of months identified as structural break points.

Assignment 6

Rolling CAPM Stability Analysis

This section discusses the results of a rolling window analysis of the parameters from the previous regression models between 2006 and 2024. The first subsection compares the rolling window analysis with the Chow test results from Assignment 5. The following three sections focus on analyzing the parameters (Alpha and Beta), the R-squared and the portfolio, respectively.

6.1 Comparisons with the Chow Test

The Rolling Window (RW) analysis revealed significant changes in parameters around three key periods: 2008, 2014-2016, and 2020-2022. These breakpoints likely correspond to major historical events: the subprime mortgage crisis, the implementation of the Affordable Care Act (ObamaCare), and the COVID-19 pandemic, respectively.

6.1.1 2008

All regression models had a break date signaled by the RW analysis around the end of 2008. This is coherent with the results of the Chow test, that indicated structural breaks for all stocks in those periods, with the exception of Labcorp Holdings. With regard to Labcorp Holdings, the only visible change of its parameter around that date is a sharp drop of its Beta, which quickly bounced back to its previous level. Apparently the Chow test didn't regard this as a significant parameter change.

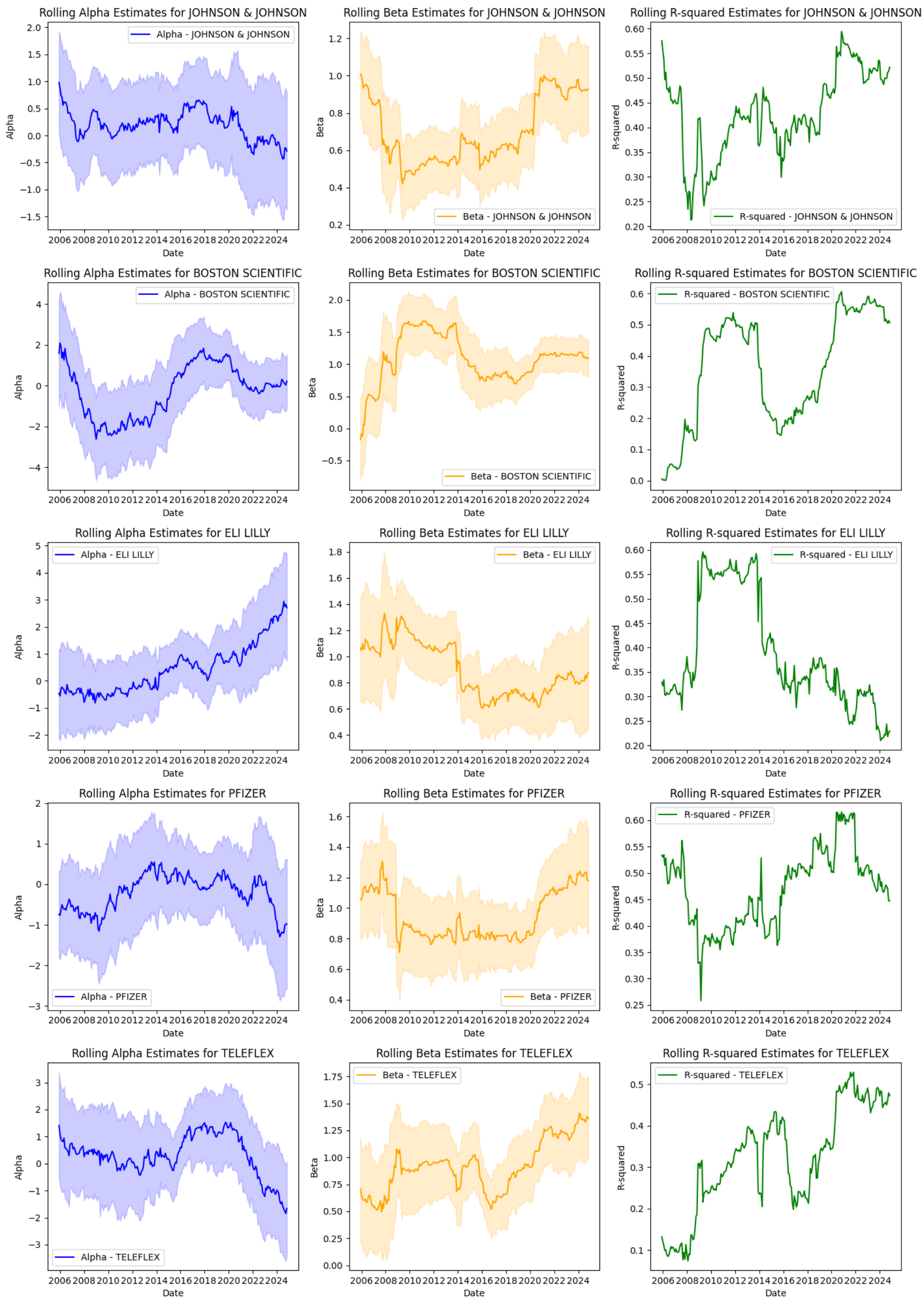

Fig. 6.1: Rolling quantities computed an all equities as an alternative of the Chow Test in search for structural breaks (1).

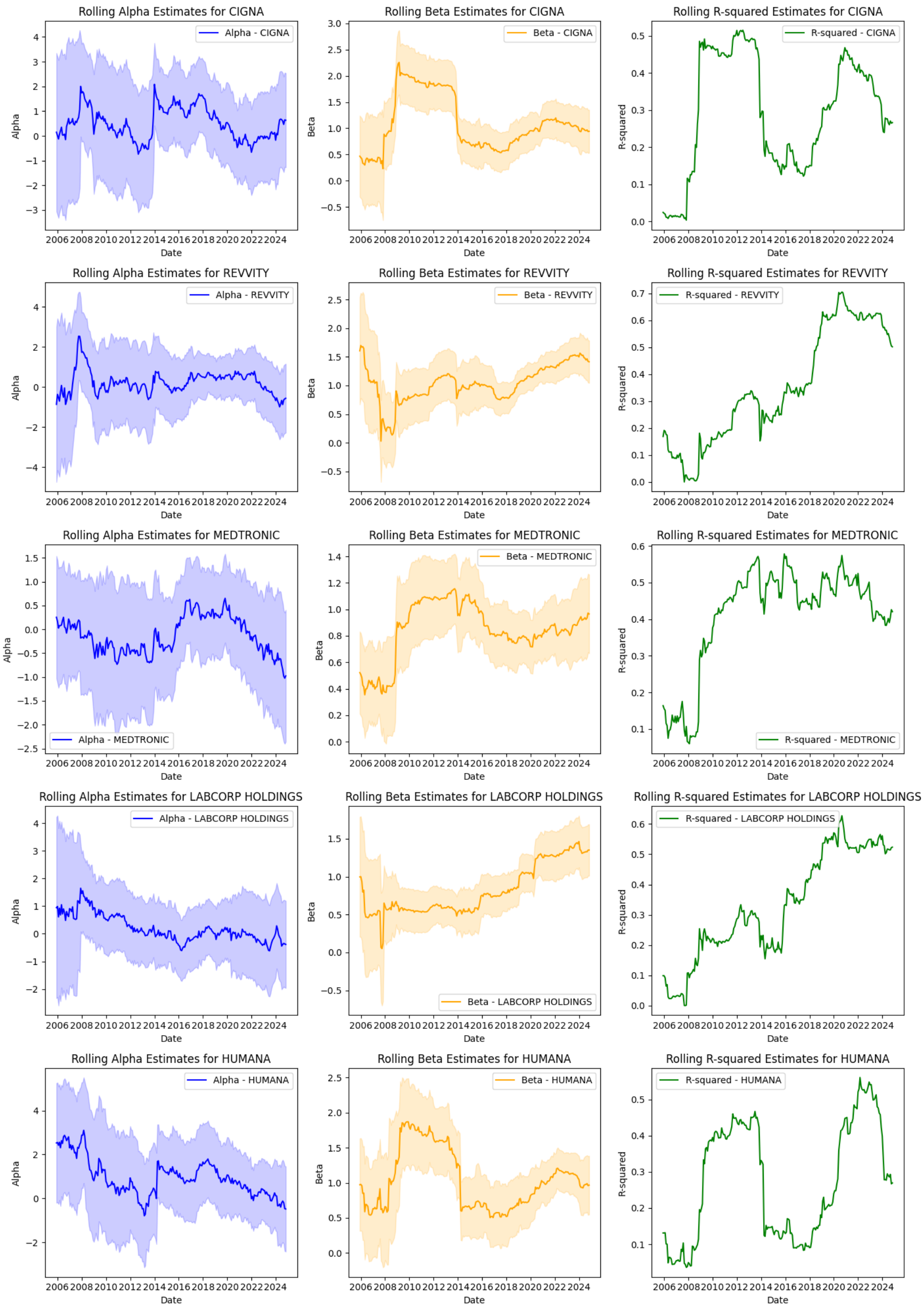

Fig. 6.2: Rolling quantities computed an all equities as an alternative of the Chow Test in search for structural breaks (2).

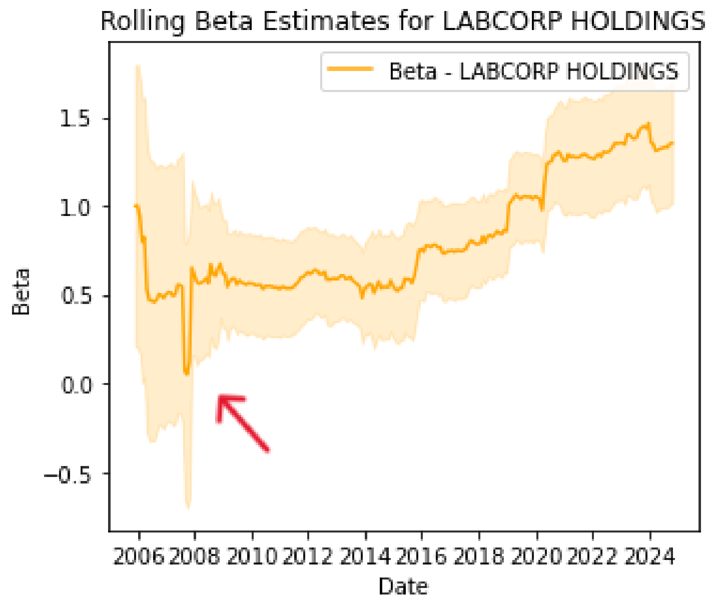

Fig. 6.3: Sharp drop of Beta parameter in Labcorp Holdings.

6.1.2 2014-2016

In these dates the results of the Chow Test and of the RW analysis are not always linked. The RW analysis detected parameter changes for 6 stocks in this period: Johnson & Johnson, Teleflex, Eli Lilly, Humana, Boston Scientific and Cigna, the last two being displayed in Figure 6.4. The Chow test, however, failed to detect structural shifts in the parameters of Boston Scientific and Cigna, despite them visually indicating important changes.

6.1.3 2020-2022

In this period the RW analysis detected parameter changes for 5 stocks in this period: Johnson & Johnson, Teleflex, Eli Lilly, Labcorp and Pfizer. Although a few other stocks exhibited parameter shifts during that period, the changes were minimal and insignificant. This aligns well with the findings of the Chow test in Assignment 5.

6.2 Alpha & Beta

The RW analysis is useful not only to detect structural breaks, but it can be useful to analyze the economic meaning of the parameters too. Generally speaking, Alphas

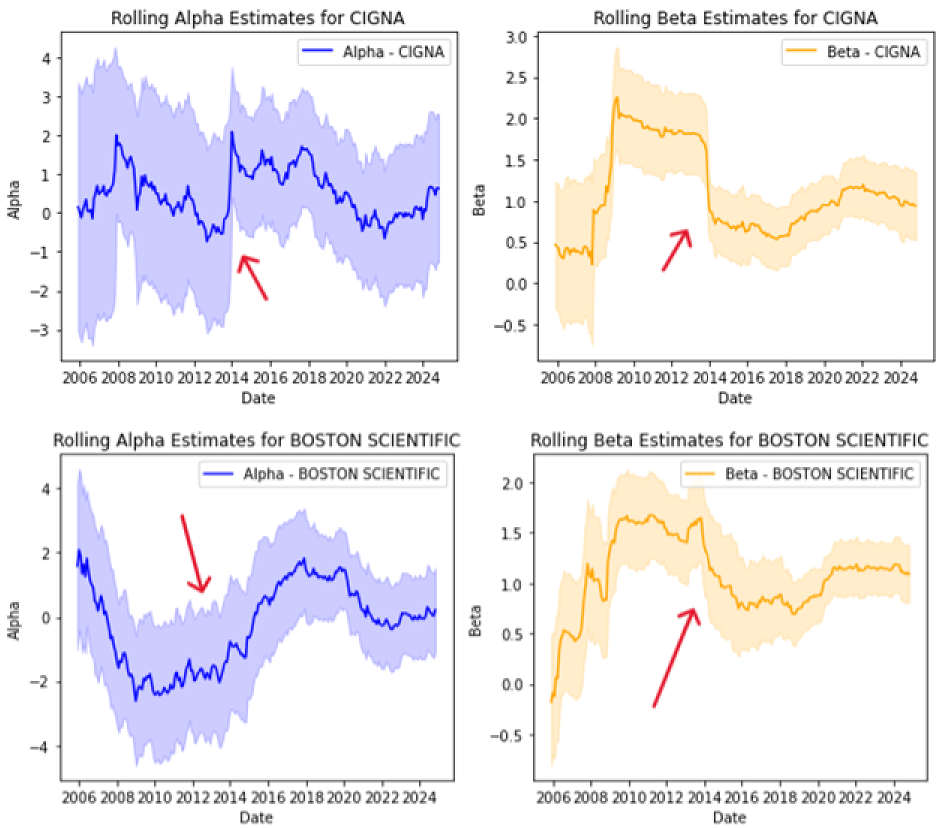

Fig. 6.4: Possible break dates identified by the rolling window.

fluctuated around 0 and Betas fluctuated around a different value for each stock. There are, however, a few exceptions:

- **Eli Lilly:** Alpha in a constant upward trend and Beta in a downward trend with some fluctuation
- **Boston Scientific:** Beta in an upward trend
- **Medtronic:** Beta in an upward trend with some fluctuation

According to the CAPM model and the Efficient Market Theory, the parameter Alpha should be null, since variables are expressed in excess returns (the risk-free rate is implicit), this is consistent with Equation 6.2.

$$r_i^e = \beta * r_m^e + \eta \tag{6.1}$$

However, when Alpha is different from zero in the regression model it can mean 2 things:

- **Alpha > 0** The stock's returns are beyond what is explained by the industry performance and risk-free rate. It overperforms in relation to its risk.

- **Alpha < 0:** The stock loses value relative to its expected risk adjusted returns. It underperforms in relation to its risk.

So, when a stock has its Alpha constantly increasing it may be a signal of the existence of positive external variables that the market hasn't yet incorporated in the stock's price—market inefficiency. That seems to be the case of Eli Lilly. The Beta, on the other hand, may assume non-null values and still be coherent with the CAPM model. From a statistical point of view, Beta represents the sensitivity of a stock's excess returns to the industry's excess returns. Whereas from an economical point of view Beta measures the sensitivity to systematic risk (economic downturns, interest rate changes, geopolitical events). A changing Beta may be detecting fundamental changes in a company's management strategy or market dynamics such as:

- Increase/decrease of leverage

- Expanding to more risky or less risky market segments

- Mergers & Acquisitions

Therefore, Eli Lilly, Medtronic and Boston Scientific changing Betas are an invitation for further analysis of the companies' business models and management strategy.

6.3 R-squared

R-squared measures the strength of the linear relationship between a stock's excess returns and the industry's excess returns. In the CAPM framework, Beta represents systematic risk, while the error term reflects idiosyncratic risk-essentially, any unexplained variance in the model. A shifting R-squared indicates a change in the relevance of the error term in explaining the dependent variable:

- **Decreasing R-squared:** The error term becomes more significant, suggesting that unknown variables are driving the stock's returns.

- **Increasing R-squared:** Beta and industry returns explain the stock's performance more effectively, making it more predictable based on systematic factors.

Periods of low R-squared may signal market inefficiencies. Investors who identify the specific factors driving a stock's deviation from market trends can develop strong, informed investment theses. Conversely, periods of high R-squared indicate alignment with market trends, where stock-specific opportunities are limited, and broader market strategies may be more effective.

The R-squared fluctuated throughout the study period (2006-2024), with notable variations in 2008 and 2020, both periods of high systemic risk—the subprime crisis and COVID-19 pandemic. In 2008, R-squared behaved differently across stocks: it increased for some, indicating stronger ties to market trends and greater impact from the crisis, while it decreased for others, suggesting weaker market influence. In contrast, 2020 saw R-squared increase for all analyzed stocks, highlighting COVID-19 as a major systemic risk that significantly explained returns across the board.

6.4 The portfolio

The portfolio's Beta and Alpha visually signal 3 possible structural break periods, 2008, 2016 and 2020.

This result does not entirely align with the Chow test, as no structural break was identified in 2020. Despite relatively high R-squared values (ranging from 0.65 to 0.85), the data displays trends similar to stock movements, characterized by notable fluctuations. Three significant changes stand out:

- **2008:** A sharp decline followed by a quick recovery.
- **2014:** A marked drop.
- **2020:** A drastic increase.

Assignment 7

Multivariate Regression

7.1 Model Introduction and Variables

In this section, we introduce new variables derived from the Fama-French 5-Factor Model and Federal Reserve Economic Data (FRED). The Fama-French model was developed to improve the evaluation of stock returns. It includes:

1. **eMKT:** The excess market returns.

2. **Mkt-Rf:** The difference between the market returns of a portfolio (comprising stocks from NYSE, AMEX, or NASDAQ) and the risk-free interest rate (one-month Treasury bill).

3. **SMB (Small Minus Big):** A factor that captures the size effect by considering the difference in returns between small and large-cap stocks.

4. **HML (High Minus Low):** Reflecting the value effect, it measures the difference in returns between high and low book-to-market value stocks.

5. **RMW (Robust Minus Weak):** A profitability factor comparing returns of companies with robust and weak profitability.

6. **CMA (Conservative Minus Aggressive):** A factor that considers investment policies, contrasting companies with conservative versus aggressive investment approaches.

To account for the macroeconomic impact on our sector, we used four additional variables:

7. **CPI (onsumer Price Index):** Representing inflation, we calculated its logarithmic changes to reflect the inflation rate.
8. **Oil Prices:** Capturing variations in energy costs.
9. **US Industrial Production Index:** Transformed logarithmically to examine coherence with growth rate of the industrial sector.
10. **Producer Price Index (PPI) for Chemical Manufacturing:** Evaluating how sector-specific costs impact returns.

We included these macroeconomic indicators because they potentially influence costs and returns for companies in our sector.

7.2 Model Building and Refinement

Initially, we estimated a regression model using all variables. Next, we iteratively removed insignificant variables using the GETS (General-to-Specific) modeling strategy, which simplifies models while retaining predictive power. The complete model is presented in Figure 7.1.

We can clearly see that many betas are not significant, so it is better if we eliminate and rerun the model with only the significant ones. The restricted model includes eMKT, MktRf, SMB, HML, and IND_PROD. While the alpha was insignificant, it was retained for comparison purposes. For all the following analysis, we will use this model as a benchmark (see Figure 7.2).

Complete vs Restricted Model: The adjusted R^2 decreased minimally from 0.790 to 0.788, confirming that the simpler model retained explanatory power. Moreover, the test on the difference of the residuals between the restricted and the unrestricted model gives a p-value equal to 0.1695, so we can agree that the difference in prediction power is not statistically significant.

Correlation between variables: Analyzing the multicollinearity, we observed a high correlation between eMKT and Mkt-rf, as shown in Figure 7.3.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.798
Model:                            OLS   Adj. R-squared:                  0.790
Method:                 Least Squares   F-statistic:                     107.7
Date:                Wed, 27 Nov 2024   Prob (F-statistic):           1.14e-88
Time:                        19:22:39   Log-Likelihood:                -589.55
No. Observations:                 284   AIC:                             1201.
Df Residuals:                     273   BIC:                             1241.
Df Model:                          10
Covariance Type:            nonrobust
================================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
Intercept     -0.4405      0.530     -0.831      0.407      -1.484       0.603
eMKT           0.7964      0.040     19.919      0.000       0.718       0.875
MktRF          0.1243      0.042      2.986      0.003       0.042       0.206
SMB           -0.1653      0.049     -3.341      0.001      -0.263      -0.068
HML            0.0653      0.050      1.296      0.196      -0.034       0.164
RMW            0.0955      0.058      1.644      0.101      -0.019       0.210
CMA            0.1440      0.074      1.953      0.052      -0.001       0.289
logCPI         0.1415      0.411      0.344      0.731      -0.668       0.951
OIL           -0.0039      0.005     -0.721      0.471      -0.014       0.007
PPI            0.0031      0.002      1.409      0.160      -0.001       0.008
logIND_PROD   -0.1896      0.088     -2.153      0.032      -0.363      -0.016
```

Fig. 7.1: Regression Summary for model complete of all variables.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.792
Model:                            OLS   Adj. R-squared:                  0.788
Method:                 Least Squares   F-statistic:                     211.6
Date:                Wed, 27 Nov 2024   Prob (F-statistic):           1.50e-92
Time:                        19:43:42   Log-Likelihood:                -593.57
No. Observations:                 284   AIC:                             1199.
Df Residuals:                     278   BIC:                             1221.
Df Model:                           5
Covariance Type:            nonrobust
================================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
Intercept      0.1628      0.119      1.365      0.173      -0.072       0.398
eMKT           0.7892      0.040     19.963      0.000       0.711       0.867
MktRF          0.1248      0.040      3.121      0.002       0.046       0.204
SMB           -0.1777      0.046     -3.901      0.000      -0.267      -0.088
CMA            0.2144      0.058      3.696      0.000       0.100       0.329
logIND_PROD   -0.1930      0.084     -2.285      0.023      -0.359      -0.027
```

Fig. 7.2: Regression summary for restricted model.

To investigate the impact, we excluded MktRF from the model to assess whether the standard error of eMKT would change significantly. After running the adjusted model (Figure 7.4), we observed that the standard error decreased, from 0.040 to 0.028. Despite this drop, we decided to keep MktRF in the model due to its low Variance Inflation Factor (VIF) of 2.618. This is well below the common threshold of 5, which is typically used to flag potential collinearity issues. We then examined the correlations among the remaining variables and found no coefficients suggesting the need for further adjustments or checks.

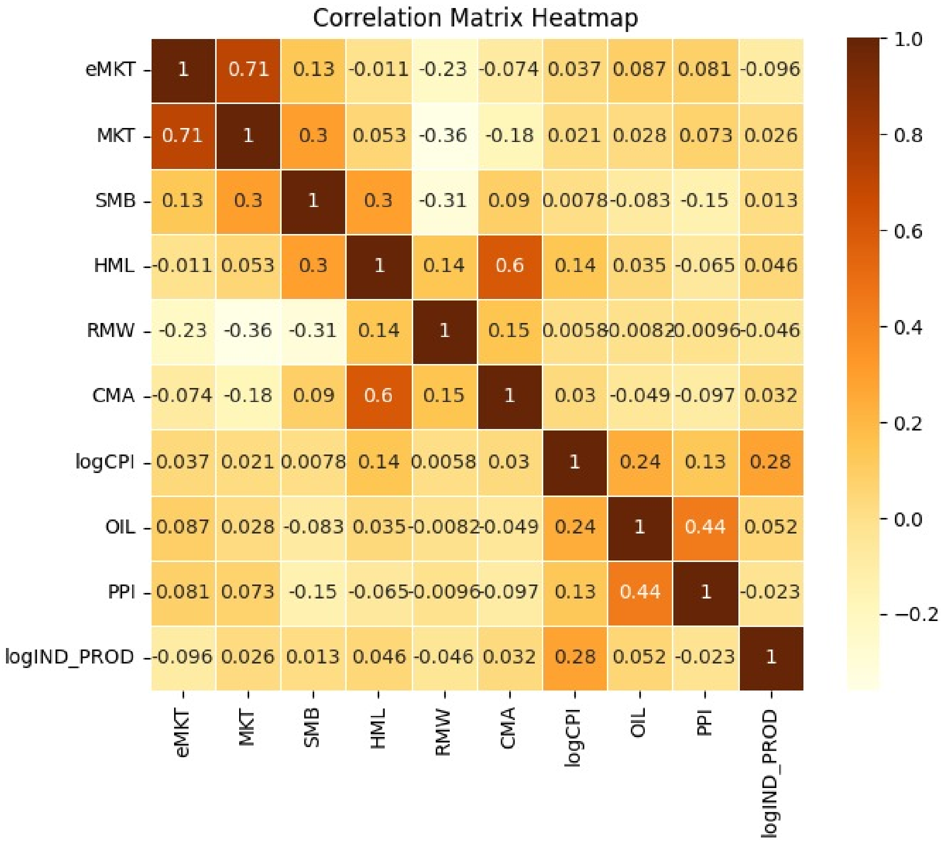

Fig. 7.3: Heatmap showing correlation between explanatory variables.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.791
Model:                            OLS   Adj. R-squared:                  0.784
Method:                 Least Squares   F-statistic:                     115.3
Date:                Wed, 27 Nov 2024   Prob (F-statistic):           8.04e-88
Time:                        19:24:17   Log-Likelihood:                -594.11
No. Observations:                 284   AIC:                             1208.
Df Residuals:                     274   BIC:                             1245.
Df Model:                           9
Covariance Type:            nonrobust
===============================================================================
                  coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------
Intercept      -0.4187      0.538     -0.779      0.437      -1.477       0.640
eMKT            0.8814      0.028     30.957      0.000       0.825       0.937
SMB            -0.1365      0.049     -2.773      0.006      -0.233      -0.040
HML             0.0971      0.050      1.945      0.053      -0.001       0.195
RMW             0.0576      0.057      1.002      0.317      -0.056       0.171
CMA             0.0832      0.072      1.158      0.248      -0.058       0.225
logCPI          0.0630      0.416      0.151      0.880      -0.756       0.882
OIL            -0.0052      0.005     -0.957      0.339      -0.016       0.006
PPI             0.0037      0.002      1.657      0.099      -0.001       0.008
logIND_PROD    -0.1517      0.088     -1.717      0.087      -0.326       0.022
```

Fig. 7.4: Regression summary for adjusted model.

7.2.1 Comparative Analysis

We estimated the CAPM using the ePortfolio as the dependent variable and eMKT as the independent variable (Figure 7.5). This model had an adjusted R^2 of 0.766, lower than the multivariate model ($R^2 = 0.788$).

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.769
Model:                            OLS   Adj. R-squared:                  0.768
Method:                 Least Squares   F-statistic:                     940.3
Date:                Wed, 27 Nov 2024   Prob (F-statistic):           8.40e-92
Time:                        19:45:50   Log-Likelihood:                -608.24
No. Observations:                 284   AIC:                             1220.
Df Residuals:                     282   BIC:                             1228.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.2162      0.123      1.755      0.080      -0.026       0.459
eMKT           0.8659      0.028     30.665      0.000       0.810       0.922
```

Fig. 7.5: Regression summary for ePortfolio.

Beta and Alpha:

- Beta: The beta for eMKT decreased from 0.8659 in the single-factor model to 0.7892 in the multivariate model, reflecting the inclusion of additional explanatory factors.
- Alpha: Both models produced insignificant alphas, with a decrease from 0.2162 (single-factor model) to 0.1628 (multivariate model).

7.2.2 Residual Analysis:

We examined the residuals of both models to assess their behavior over time and distributional properties.

Residual Variance: CAPM Residuals displayed greater variance (red line in Figure 7.6), consistent with a poorer fit, while Multivariate Model Residuals variance was lower (blue line in Figure 7.6), indicating a better fit.

Residual Distribution: CAPM Residuals followed a normal distribution but with a longer left tail, suggesting the presence of outliers (Figure 7.7).

Multivariate Model Residuals also followed a normal distribution but were more concentrated around the mean (Figure 7.8), reflecting the better explanatory power of the model.

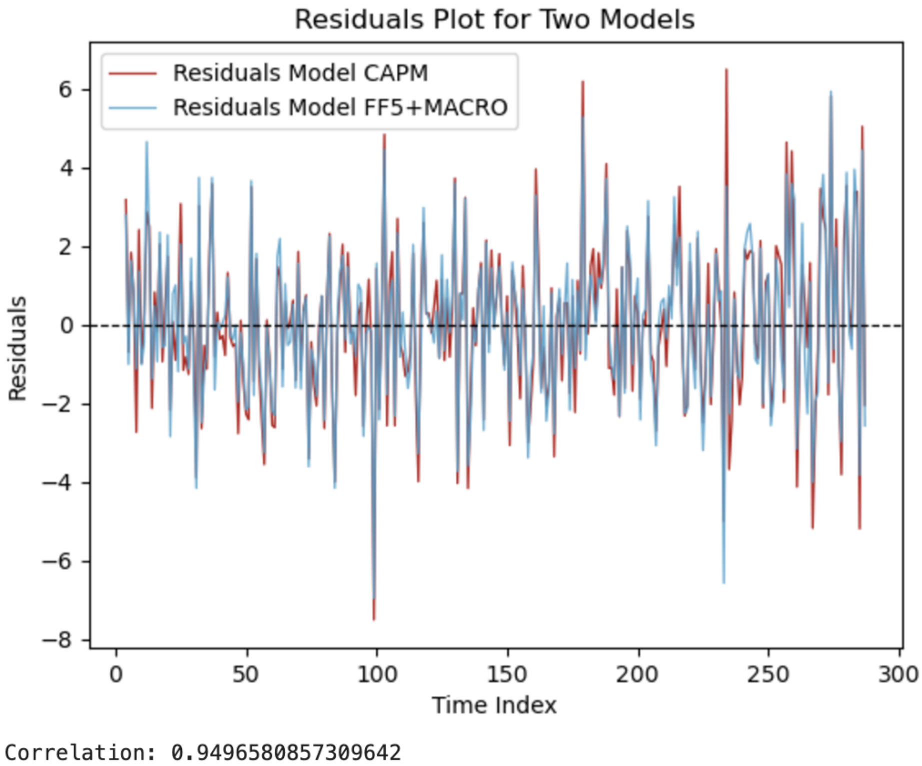

Fig. 7.6: CAPM Residuals.

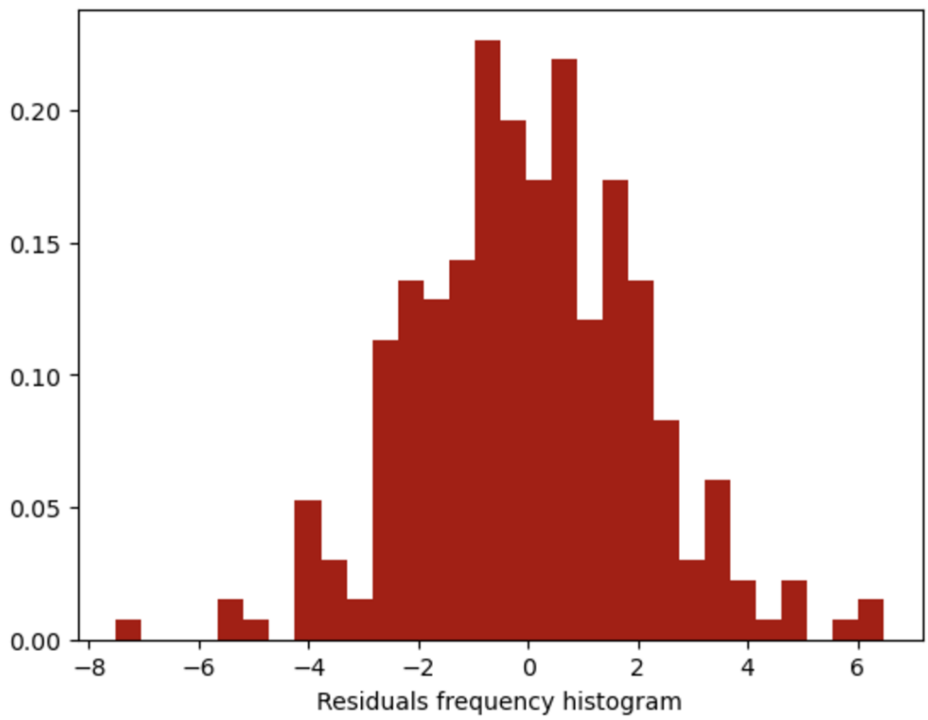

Fig. 7.7: Histogram of residuals for univariate model.

Residual Correlation: As shown in Figure 7.9, residuals from both models had low autocorrelation, indicating reliable coefficient estimation.

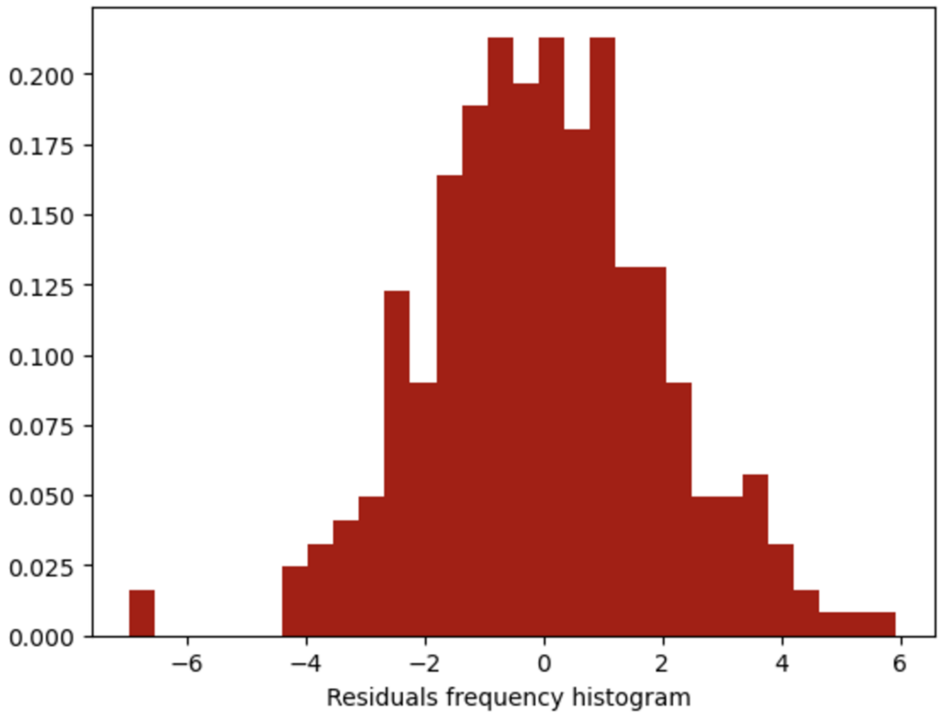

Fig. 7.8: Histogram of residuals for multivariate model.

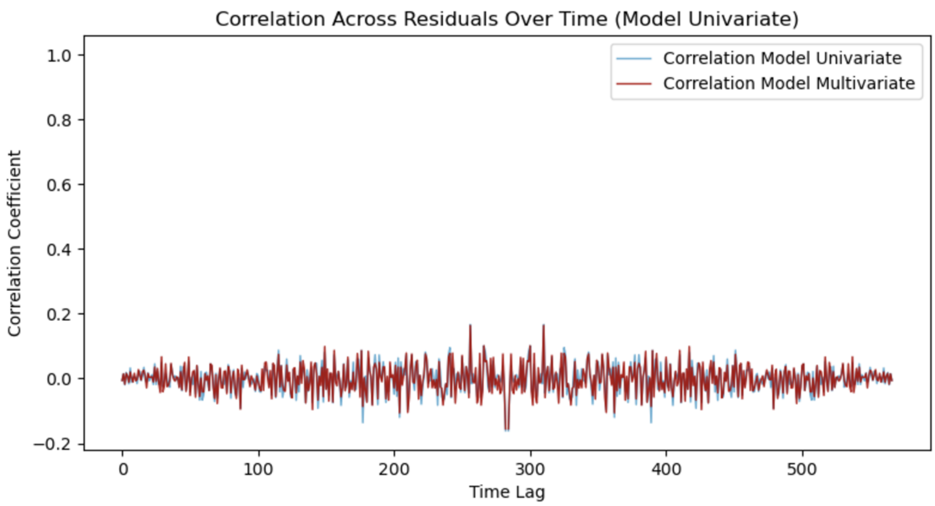

Fig. 7.9: Correlation of residuals over time.

Conclusion

By incorporating the Fama-French factors and additional macroeconomic variables into the CAPM framework, we achieved a model with improved explanatory power. The multivariate model captures sector dynamics more effectively than the single-factor CAPM, as evidenced by better residual behavior and higher adjusted R^2.

Appendix

Summary of Group Members' Contribution

Table 7.1 provides a detailed breakdown of each group member's contributions to the project. Notably, Giorgio Cottini also supervised the overall coding process and authored the accompanying LaTeX script.

Assignment	Code	Comment
Assignment 1	Enrico Paciaroni	Enrico Paciaroni
Assignment 2	Luigi Babiski	Martina Arrighini
Assignment 3	Giorgio Cottini	Martina Arrighini
Assignment 4	Martina Arrighini	Martina Arrighini
Assignment 5	Giorgio Cottini	Giorgio Cottini
Assignment 6	Luigi Babiski	Luigi Babiski
Assignment 7	Enrico Paciaroni	Enrico Paciaroni

Table 7.1: Contribution of Group Members to Assignments