Group Assignment

Financial Econometrics (FIN 36182)

Fall 2024

Group Members:

Baldoni Chiara
Pinna Paola
Torella Marta

After conducting a minimal EDA to check for missing values and duplicates, we converted the returns to decimal form. Consequently, all results in the following analysis are presented in decimal format.

TASK 1

1) – 3) In the below table, we report the arithmetic mean of the returns for each of the five industries ("Cnsmr", "Manuf", "HiTec", "Hlth", "Other") over the entire sample, the standard deviation of the returns for each of the five industries over the entire sample, and the Sharpe ratio of each industry.

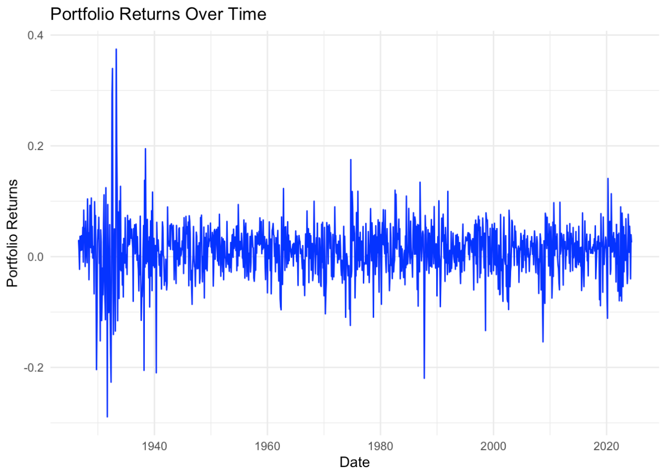
	Mean_Returns <dbl>	Std_Returns <dbl>	Sharpe_Ratio <dbl>
Cnsmr	0.1215	0.1828	0.6647
Manuf	0.1152	0.1905	0.6047
HiTec	0.1203	0.1935	0.6217
Hlth	0.1288	0.1910	0.6743
Other	0.1104	0.2211	0.4993

4) Based on the above results, technology stocks (represented by the "HiTec" industry) have a Sharpe ratio of 0.6217, indicating better risk-adjusted returns compared to the manufacturing (0.6047) and other sectors (0.4993). However, they do not outperform the consumer (0.6647) and health industries (0.6743), which exhibit higher Sharpe ratios. Therefore, while technology stocks offer relatively good risk-adjusted returns, they are not the best performers overall, as both the consumer and health sectors deliver better risk-adjusted outcomes.

5) The below correlation table shows strong positive correlations across the five industries, indicating that their returns tend to move in the same direction. The highest correlations are observed between the Manufacturing and Other sectors (0.8924) and between the Consumer and Other sectors (0.8716), suggesting a particularly close relationship in returns among these industries. Conversely, Technology (HiTec) shows the lowest average correlation with other sectors, with values ranging from 0.7072 (Health) to 0.8164 (Consumer), indicating it is slightly less correlated with the rest. Overall, the high correlations imply limited diversification benefits when combining these industries, as they are likely to experience similar return patterns.

	Cnsmr <dbl>	Manuf <dbl>	HiTec <dbl>	Hlth <dbl>	Other <dbl>
Cnsmr	1.0000	0.8670	0.8164	0.7739	0.8716
Manuf	0.8670	1.0000	0.7996	0.7427	0.8924
HiTec	0.8164	0.7996	1.0000	0.7072	0.7932
Hlth	0.7739	0.7427	0.7072	1.0000	0.7371
Other	0.8716	0.8924	0.7932	0.7371	1.0000

6) After constructing a time series of the simple, non-cumulative returns of the provided data, where capital is allocated equally across the first four industries (excluding Other), as displayed by the below plot of the portfolio return over time, we report in a table the arithmetic mean, standard deviation and Sharpe ratio of the diversified portfolio.



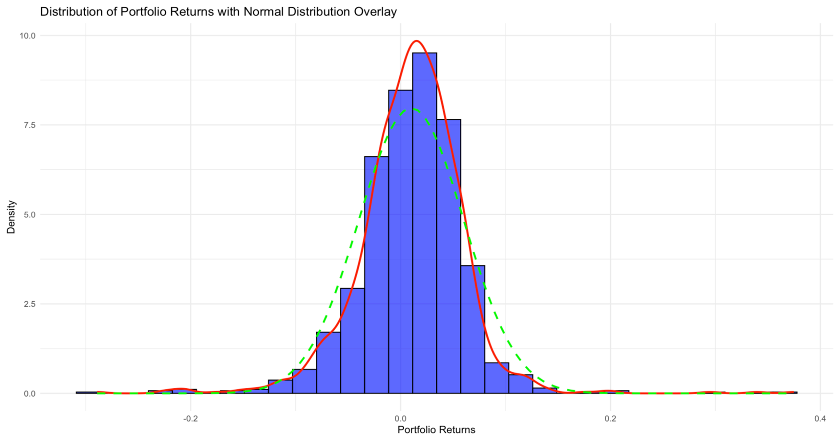
Metric <chr>	Value <dbl>
Annualized Mean Return	0.1214
Annualized Standard Deviation	0.1734
Sharpe Ratio	0.7004

As showed by the table above, the diversified portfolio, demonstrates moderate risk-adjusted performance over time. The annualized mean return of 0.1214 suggests a steady average gain, while the annualized standard deviation of 0.1734 indicates a level of volatility that’s relatively contained but could still expose the portfolio to fluctuations in specific periods, as evidenced by the peaks in the time series plot. The Sharpe ratio of 0.7004, while not exceptionally high, indicates that the returns provide a reasonable compensation for the risk taken. This diversified approach likely mitigates some industry-specific risk, offering a balanced trade-off between return and volatility, although with limited excess return relative to the risk-free benchmark.

TASK 2

1) - 2) In the below table we report the kurtosis and skeweness for the market portfolio we obtained from point 1.6); moreover, to better highlight the differences between the distribution of the returns R_m in the market portfolio and a normal distribution, the below plot shows the distribution of portfolio returns (blue histogram with red density curve) overlaid with a normal distribution (green dashed line).

Metric <chr>	Value <dbl>
Kurtosis	10.2856
Skewness	0.0237

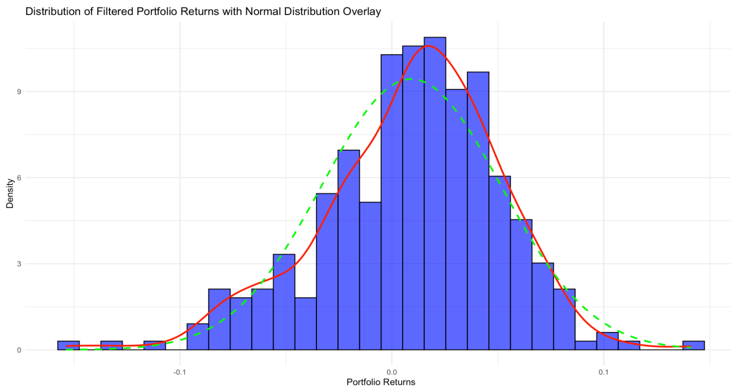


The market portfolio displays a kurtosis of 10.2856, which is significantly higher than the kurtosis of a normal distribution, which is 3. Indeed, this high value indicates that the distribution of R_m is leptokurtic, meaning that it has heavier tails, and a greater likelihood of extreme returns compared to a normal distribution. This element is also evident in the plot, where the observed returns display more "pronounced" tails than the overlaid normal distribution, with subtle "peaks" that highlight the distribution's heavier tails. The skewness of R_m, on the other hand, is 0.0237, which is very close to zero, suggesting that the distribution is nearly symmetric; this symmetry is also reflected in the histogram and density plot, where the returns are centred around the mean without significant skew to either side. Overall, while the distribution of R_m is nearly symmetric like a normal distribution, the high kurtosis indicates a much higher probability of extreme values, highlighting the risk of large deviations from the mean in this portfolio.

3) In the table below, we report again the kurtosis and skewedness for the market portfolio but eliminating the first 70 years of data; we also show the plot of the

distribution of portfolio returns overlaid with a normal distribution with data starting from 1997-06.

Metric <chr>	Value <dbl>
Kurtosis	3.7380
Skewness	-0.5123



For the selected sample, the kurtosis of 3.7380, close to the normal distribution's kurtosis of 3, indicates that the sample has only slightly fat tails, meaning fewer extreme values than before, resulting in a distribution closer to normality. The skewness of -0.5123 suggests a modest left skew, implying that negative returns are somewhat more frequent or extreme than positive returns. Overall, recent data shows a shift toward a more normalized distribution with a slight tendency toward negative returns.

4) The covariances between the first four industries' returns and the portfolio returns R_m are reported in the following table.

Industry <chr>	Covariance_with_Rm <dbl>
Cnsmr	0.0025
Manuf	0.0026
HiTec	0.0025
Hlth	0.0024

The positive covariance values imply that all industries move in the same direction as the market. However, the relatively small values suggest that, while affected by market movements, they each maintain some degree of independence.

5) – 6) After computing the beta values for the first four industries for the full sample (Beta_Full_Sample), we repeated the computation for the sample starting from June 1997 (Beta_Filtered). Both results are reported in the following table.

Industry <chr>	Beta_Full_Sample <dbl>	Beta_Filtered <dbl>
Cnsmr	1.00	0.9236
Manuf	1.04	0.9533
HiTec	1.00	1.3305
Hlth	0.96	0.7926

The comparison of beta values between the full sample and the post-June 1997 period reveals meaningful shifts in the market sensitivity of various industries. In CAPM, beta measures a stock's sensitivity to market movements, with a beta of 1 indicating that the stock's returns tend to move in line with the market. For Consumer stocks, beta declines slightly from 1.00 in the full sample to 0.9236 in the filtered period, suggesting that Consumer stocks have become less responsive to overall market fluctuations. Similarly, Manufacturing shows a decrease in beta from 1.04 to 0.9533, indicating that its returns have become less volatile relative to the market in recent years. In contrast, Technology stocks exhibit a substantial increase in beta, rising from 1.00 to 1.3305, implying that the sector has become significantly more reactive to market movements since 1997, potentially due to the sector's rapid growth and increased market integration. Lastly, Health stocks also see a reduction in beta, dropping from 0.96 to 0.7926, indicating an even lower sensitivity to market shifts, making it closer to a defensive asset under CAPM. Overall, these beta changes suggest that, in the post-1997 period, Technology has become more volatile and risk-exposed to the market, while Consumer, Manufacturing, and Health sectors have shown a trend toward lower market sensitivity, indicating a shift toward reduced market dependence.

7) After computing the Jensen's alpha values assuming a risk-free rate of 5% for the first four industries, we report the results in the following table in percentage terms.

	Industry <chr>	Alpha <dbl>
Cnsmr	Cnsmr	0.0051
Manuf	Manuf	−0.9107
HiTec	HiTec	−0.1149
Hlth	Hlth	1.0209

The risk-free rate represents the return you would expect from an investment with no risk; hence, by using this 5% risk-free rate we're setting a relatively high baseline for expected returns, which means that for alphas to be positive, the industries should outperform a model predicted return that starts from this 5% base. However, notice that the CAPM model assumes that the alpha should be zero, as it predicts that returns are entirely explained by systematic risk (beta) alone. Therefore, any deviation from zero alpha suggests that the CAPM may not fully account for all the risk factors influencing returns in these industries
The Consumer's industry small positive alpha (0.0051%) tells us that it has slightly outperformed its CAPM-predicted return, achieving excess returns on a risk adjusted basis. On the other hand, the Manufacturing industry shows a significant negative alpha (-0.9107%) suggesting that it underperformed the CAPM expectations, probably due to weak market conditions. The negative alpha for High-Tech industry (-0.1149%) also indicates underperformance but less severe compared to Manufacturing; this could be due to volatility, as fluctuating returns may lead to lower-than-expected results in some periods. Finally, a high positive alpha for the Healthcare industry (1.0209%) suggests that it significantly outperformed its CAPM-predicted return, probably due to strong demand stability and favourable market conditions.

TASK 3

In the following sections, we use the lm() function to perform the linear regressions of $R_m(t)$ on an intercept (included by default in R) and respectively on $R_m(t-1)$, $(P_m(t-1) / P_m(t-13))-1$, and $abs(R_m(t))$, on both the full sample and the sample starting from 1997-06.
Each time, in the model summary, we report the estimates, the standard error, the t-statistics, and the p-values for both the intercept and the predictor. In particular, the t-statistics tests whether the coefficients are significantly different from zero, indicating a meaningful relationship with the dependent variable. For large samples, a t-statistic of about 2 (1.96) corresponds to a p-value of 0.05, which is the common threshold for significance. In regression, the null hypothesis for the intercept is that it is equal to zero, implying that, if true, there would be no baseline effect when the predictor is zero; while for the predictor, is that its coefficient is zero, indicating no relationship.
Due to constraints on space, the plots illustrating the linear regression analysis are

not presented within this paper. However, they are accessible in the project's GitHub repository, as referenced in the appendix

1) To examine whether past returns influence current returns, we regress $R_m(t)$ on an intercept and on $R_m(t - 1)$, the lagged version of the portfolio returns labelled as Rm_lag in the summary below.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.009287   0.001486   6.249 5.76e-10 ***
Rm_lag      0.080911   0.029101   2.780  0.00552 **
```

The regression results show a small but statistically significant positive relationship between lagged returns $R_m(t - 1)$ and current portfolio returns, with a coefficient of 0.0809, indicating that a 1-unit increase in past returns leads to a 0.0809-unit increase in current returns. This relationship is evidenced by a low p-value (0.0056) and a t-statistic of 2.780 for lagged returns, indicating that the coefficient is 2.78 standard errors away from zero, which allows us to reject the null hypothesis of no relationship. The intercept is also significant, with an estimated return of about 0.93% when $R_m(t - 1)$ is zero, accompanied by a very high t-statistics of 6.249.

2) To examine whether long-term price changes impact current returns, we regress $R_m(t)$ on an intercept and on $(P_m(t-1) / P_m(t-13))-1$, the percentage change in price over the past year (from 12 months ago to one month ago) labelled as Pm_ratio in the summary below. In particular, we generate a price series $P_m(t)$ from the portfolio returns data, starting with an initial value of 1, and calculate each subsequent price iteratively based on the relationship $R_m(t) = (P_m(t) / P_m(t-1))-1$. We then create two lagged price variables, $P_m(t-1)$ *and* $P_m(t-13)$, and use these to calculate Pm_ratio.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.009275   0.001764   5.259 1.72e-07 ***
Pm_ratio    0.005439   0.007567   0.719    0.472
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.05025 on 1161 degrees of freedom
  (13 observations deleted due to missingness)
Multiple R-squared:  0.0004448, Adjusted R-squared:  -0.0004161
F-statistic: 0.5167 on 1 and 1161 DF,  p-value: 0.4724
```

The regression results show that the intercept, estimated at 0.0093, is statistically significant with a high t-statistics of 5.259, suggesting that the expected portfolio return is around 0.93% when the predictor $(P_m(t-1) / P_m(t-13))-1$, representing the percentage change in price over the past year, is zero. However, the coefficient for this predictor, 0.0054, has a high p-value (0.4720) and a low t-statistic of 0.719, providing weak evidence against the null hypothesis of no relationship, and suggesting, in turn, that long-term price changes have no meaningful impact on current returns.

3) The t-statistics reported by lm() in the above regression may not be very reliable due to the low R-squared value of the model, which is very close to zero (0.0004). Such a low R-squared value indicates that the model explains almost none of the variability, meaning that the predictor $(P_m(t-1) / P_m(t-13))-1$ contributes very little to predicting the portfolio returns $R_m(t)$.
In this case, the standard errors may not accurately capture data variability, potentially leading to misleading t-statistics. Consequently, the t-statistics should be interpreted with caution, given the model's lack of explanatory power. Additionally, the high p-value for the predictor (0.4720) indicates that its coefficient is not statistically significant, further supporting the idea that random variation plays a substantial role in the data.

4) To investigate whether the size of a return impacts the return itself, we regress $R_m(t)$ on an intercept and $|R_m(t)|$, the absolute version of the portfolio return, labelled as abs_Rm in the summary below. Since $|R_m(t)|$ and $R_m(t)$ are derived from the same value, this regression may be redundant, as knowing the absolute value already reveals information about the return.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.003313   0.002105   1.573    0.116
abs_Rm      0.183720   0.041245   4.454 9.22e-06 ***
```

The regression results show that the absolute returns $|R_m(t)|$ have a statistically significant positive relationship with portfolio returns. With a coefficient of 0.1837, and a t-statistic of 4.454, this suggests that larger return magnitudes are associated with higher return values. The high t-statistic indicates that the coefficient for $|R_m(t)|$ is significantly different from zero, providing strong evidence of this relationship. The intercept, estimated at 0.0033, is not statistically significant, given the high p-value (0.116) a t-statistic of 1.573, indicating that when the absolute return is zero, the expected return is close to zero and lacks statistical relevance. However, since $|R_m(t)|$ and $R_m(t)$ refer to the same time period, this analysis does not offer practical insights into the predictive behaviour of returns.

5) We repeated the regression in point 3.1) on the data from June 1997, excluding the first 70 years.

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.009114   0.002407   3.787 0.000182 ***
Rm_lag      -0.008882   0.055668  -0.160 0.873335
```

The new regression results differ notably from the original findings. Here, the coefficient for the lagged returns $R_m(t-1)$ is -0.0089, indicating a slight negative relationship between lagged and current portfolio returns. However, this coefficient is not statistically significant, as indicated by the very low t-statistic of -0.160 and a high p-value (0.8733), suggesting that the data from June 1997 onward has no meaningful influence of past returns on current returns. This contrasts with the previous model, where the lagged returns showed a small but statistically significant positive effect on current returns. The intercept, estimated at 0.0091 with a high t-statistic of 3.787, remains statistically significant, suggesting a baseline return of around 0.91% when $R_m(t-1)$ is zero, similar to before.

6) We repeated the regression in point 3.2) on the data from June 1997.

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.151e-03  2.928e-03   3.125  0.00194 **
Pm_ratio    -7.716e-07  1.501e-02   0.000  0.99996
```

The new regression results are consistent with the original findings, confirming that whether the portfolio's price has increased or decreased over the past year does not meaningfully influence current returns. The coefficient for the percentage change in price from 12 months ago to one month ago is -7.716×10^{-7}, indicating a meaningless relationship with current portfolio returns, as underscored by the t-statistic of 0.000. This aligns with the previous model's conclusion with an even stronger indication of insignificance from June 1997. The intercept is estimated at 0.0092 with a t-statistic of 3.125 and a low p-value (0.0019), indicating a statistically significant baseline return of approximately 0.92% when the percentage change in price is zero, which is consistent with the full sample result.

7) We repeated the regression in point 3.4) on the data from June 1997.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.007406   0.003891   1.903   0.0579 .
abs_Rm      0.050668   0.090091   0.562   0.5742
```

The new regression results differ from the original findings, indicating that the positive relationship between the magnitude of the return $|R_m(t)|$ and the return itself observed before does not hold in the data from June 1997 onward. Here, the coefficient for $|R_m(t)|$, 0.0507, lacks statistical significance, as shown by the low t-statistic of 0.562 and a high p-value (0.5742), implying that any apparent association between the magnitude and value of the return is likely due to random variation. The intercept, estimated at 0.0074 with a t-statistic of 1.903 suggests a baseline return of around 0.74% when the magnitude of the return is zero. However, this result is only weakly significant.