Financial Assignment

Part 1

The first part of your task:

1. Read the data: excel file “First Assignment dataset”;
2. From S1-S10 and the risk-free rate data, create excess returns for the ten portfolios. Call these excess returns XS1-XS10;
3. Inspect the returns for each portfolio and comment on any exceptional features;
4. Compute summary statistics for the 10 excess return series. Comment on the variation in statistics across the portfolios and the implications of this variation.

Summary stats table:

	Mean	Std	Min	P25	Median	P75	Max	Skewness	Kurtosis	Sharpe_ann
XS1	1.09%	5.41%	-18.36%	-2.07%	0.97%	3.79%	27.25%	0.49	3.50	0.70
XS2	1.05%	5.53%	-21.27%	-1.75%	1.13%	3.99%	26.82%	-0.04	3.01	0.66
XS3	0.87%	5.37%	-24.59%	-1.72%	1.07%	4.14%	29.28%	-0.27	4.14	0.56
XS4	0.79%	5.55%	-22.68%	-2.07%	0.93%	3.74%	25.78%	-0.20	3.08	0.50
XS5	0.72%	5.49%	-23.28%	-1.91%	1.01%	3.63%	23.14%	-0.29	2.41	0.46
XS6	0.76%	5.46%	-24.31%	-2.01%	1.15%	3.64%	23.18%	-0.52	2.66	0.48
XS7	0.67%	5.45%	-24.74%	-1.93%	1.02%	3.61%	22.06%	-0.77	3.27	0.43
XS8	0.66%	5.45%	-26.24%	-1.94%	1.37%	3.90%	19.42%	-1.02	3.63	0.42
XS9	0.67%	5.53%	-29.38%	-2.23%	1.39%	4.06%	15.48%	-0.94	3.11	0.42
XS10	0.51%	4.55%	-27.70%	-1.80%	0.98%	3.32%	13.55%	-1.04	4.40	0.38

The table reports the main descriptive statistics for the monthly excess returns (XS1–XS10) of the ten size-sorted portfolios, from the smallest (XS1) to the largest (XS10).
Each column summarizes key features of the return distribution.

- **Mean** represents the average monthly excess return. It indicates the typical risk-adjusted performance of each portfolio. In most equity datasets, smaller portfolios (XS1–XS3) tend to show higher average returns, reflecting the well-known size premium.

- **Standard deviation (Std)** measures the volatility of returns. Higher values imply greater risk. We typically observe that smaller portfolios also have higher volatility, consistent with their higher expected returns.
- **Quantiles (P25, Median, P75)**, together with **Min** and **Max**, show the dispersion and asymmetry of returns. Wider gaps between percentiles suggest more variability or extreme movements in the tails of the distribution.
- **Skewness** captures the asymmetry of the return distribution.
  - A **negative** skewness means that extreme negative returns occur more often than extreme positive ones (left-tailed).
  - A **positive** skewness indicates more frequent large gains.
    For equity portfolios, a slightly negative skewness is common due to downside risk.
- **Kurtosis** measures the "tailedness" or the frequency of extreme outcomes compared to a normal distribution (which has kurtosis = 3).
  - Values **above 3** indicate fat tails — more extreme observations than expected under normality.
  - Values **close to 3** suggest a roughly normal distribution.
    In financial returns, high kurtosis (leptokurtic behavior) is typical, reflecting occasional large shocks.
- **Sharpe ratio** expresses the return per unit of risk:
  Sharpe = $E[R-Rf] / \sigma$. Higher Sharpe ratios imply better risk-adjusted performance. Usually, large-cap portfolios (XS8–XS10) exhibit lower Sharpe ratios than small-cap ones because their returns are smoother but less pronounced.

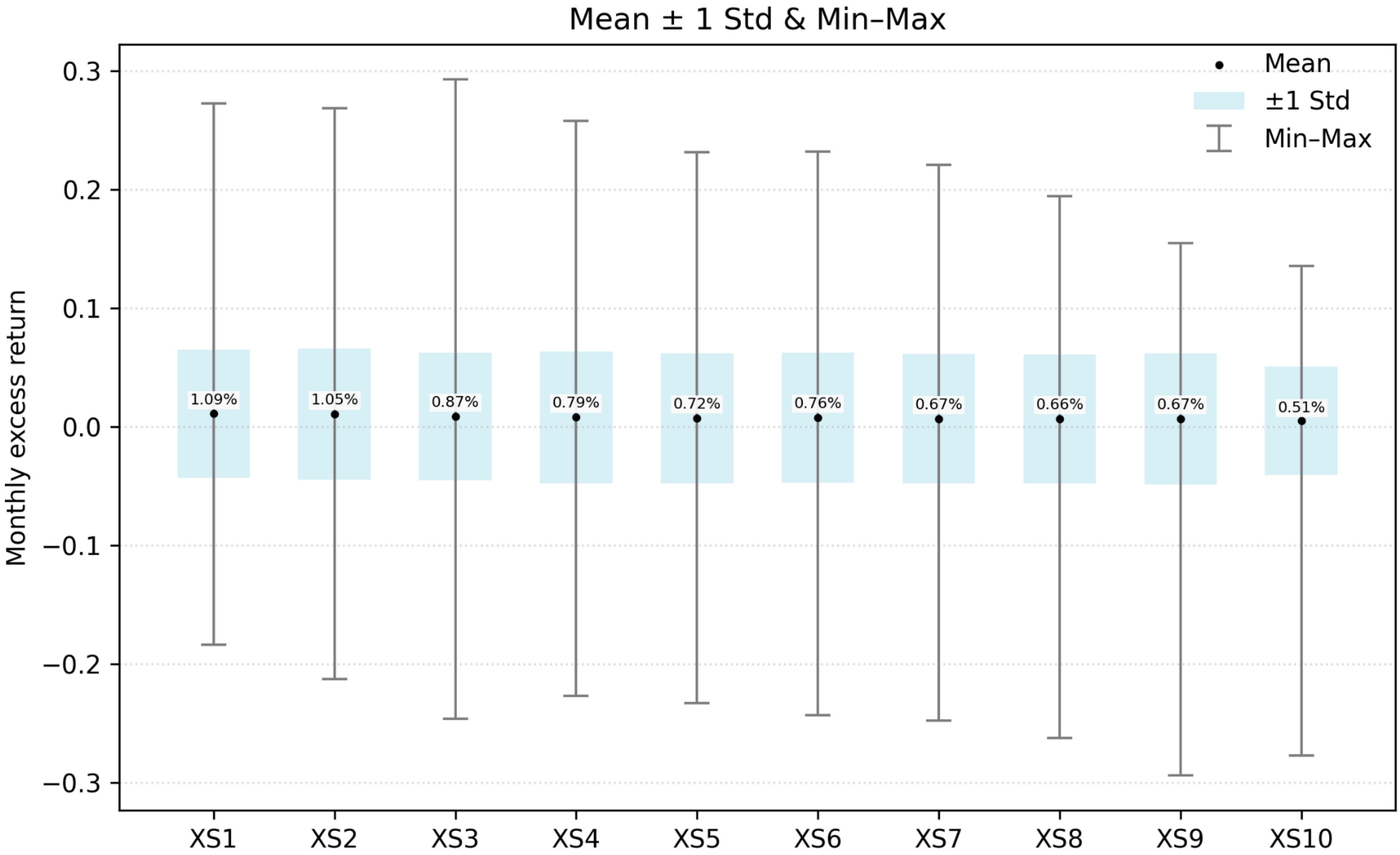
Average excess returns **(Mean)** and **Standard deviations** confirm the decreasing trend in both return and risk as firm size increases.

Skewness is mostly negative for larger portfolios, suggesting a higher likelihood of large negative returns, while smaller portfolios show slightly positive or near-zero skewness.

Kurtosis values are all above 3, indicating *fat tails* and the presence of extreme observations relative to the normal distribution.

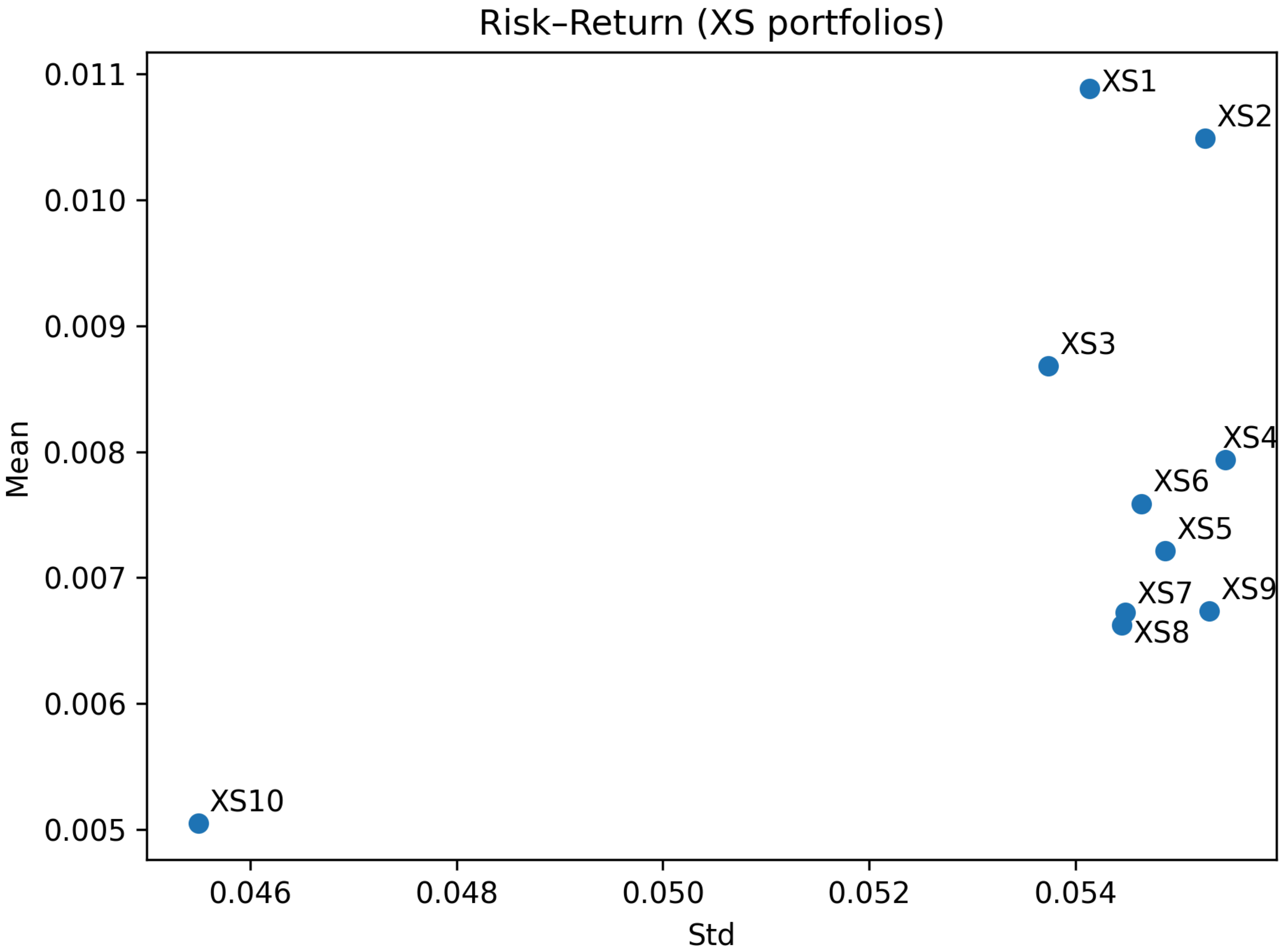
Finally, the **Sharpe ratio** declines from XS1 (0.70) to XS10 (0.38), highlighting that small-cap portfolios deliver higher risk-adjusted returns despite their greater volatility..

Mean +- 1 std & Min-Max:



The figure provides a visual summary of the distribution of monthly excess returns across the ten size portfolios (XS1–XS10).
For each portfolio, the **blue bar** represents the range of one standard deviation above and below the mean, while the **gray whiskers** indicate the **minimum and maximum** observed values.
The **black dot** marks the average excess return, and the small **label above it** reports its value in percentage terms.
The figure clearly illustrates the size effect: smaller portfolios (XS1–XS3) exhibit higher mean returns and greater volatility, while larger portfolios (XS8–XS10) show lower but more stable excess returns.

Risk-Return:



This scatter plot shows the relationship between **risk** (standard deviation of excess returns) and **average return** (mean) for the ten size-sorted portfolios.

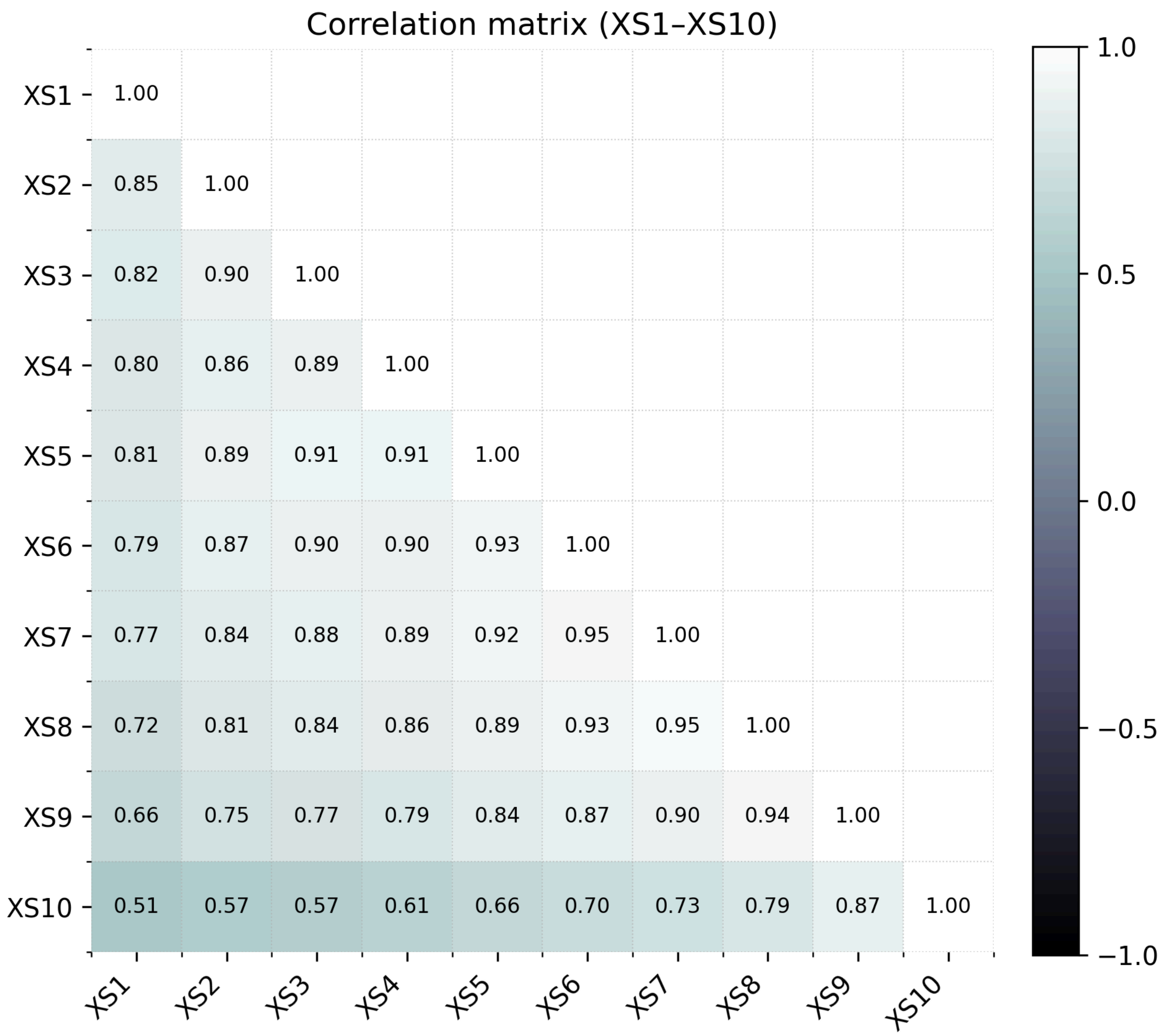
Each point represents one portfolio (XS1–XS10), moving from the smallest to the largest firms.

The upward slope indicates the typical risk–return trade-off: portfolios with higher volatility (smaller-cap stocks such as XS1–XS3) also exhibit higher average excess returns.

Notably, XS10 stands out from the others, lying well below the general trend — it shows significantly lower risk but also a substantially smaller mean excess return, highlighting the weaker performance of the largest-cap stocks.

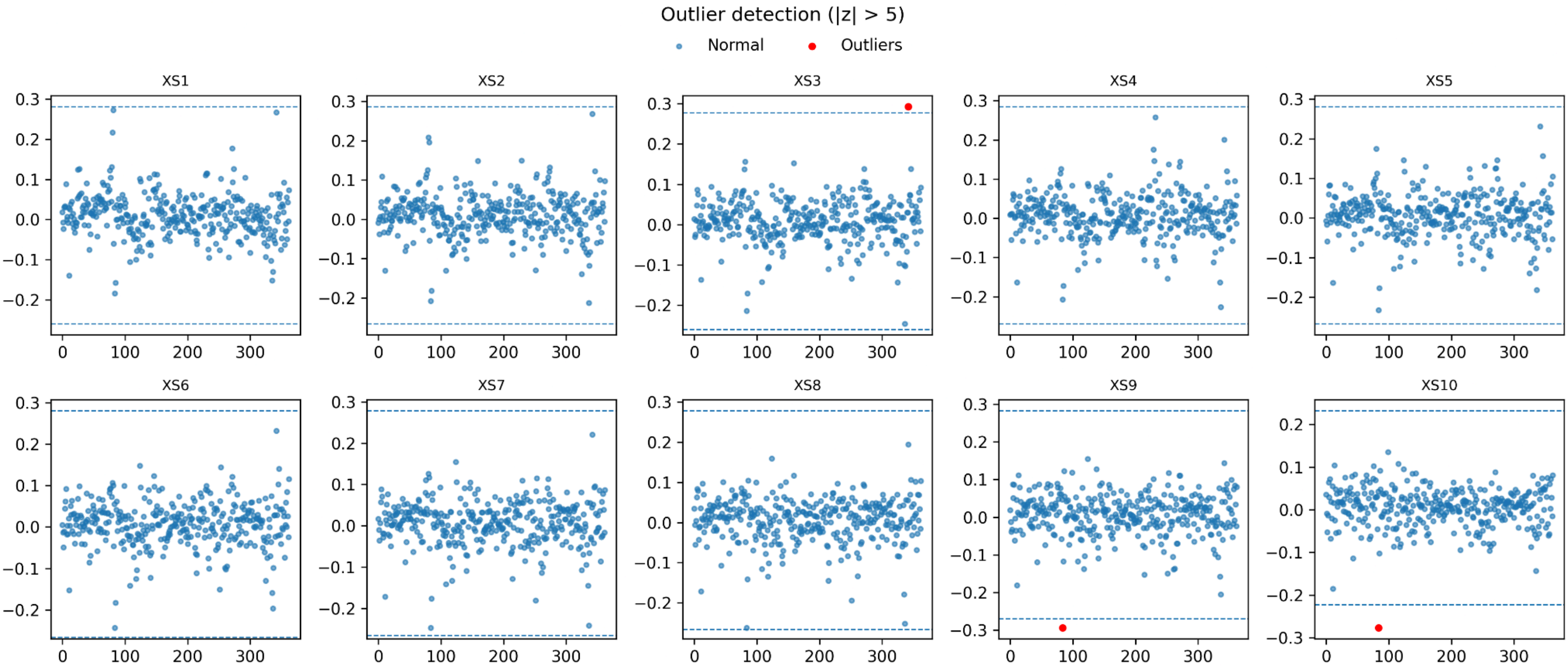
This pattern visually confirms the presence of the size effect in the UK market data.

Correlation matrix:



The heatmap shows the pairwise correlations among the monthly excess returns of the ten size portfolios.
All correlations are **strongly positive**, confirming that portfolios formed on firm size tend to move together, reflecting common market-wide risk factors.
Correlations are highest among portfolios with **similar size ranks** (e.g., XS4–XS7), and they gradually decrease as the difference in size increases.
Notably, XS10, which contains the largest-cap stocks, exhibits the lowest correlation with the smaller portfolios, indicating that large firms are less sensitive to the same shocks affecting small-cap segments of the market.

Outliers:



This grid shows the detection of extreme monthly excess returns across the ten portfolios (XS1–XS10). Each blue dot represents a single monthly observation, while red dots identify outliers, defined as returns exceeding five standard deviations from the mean (|z| > 5).
The dashed horizontal lines mark the upper and lower cut-off thresholds.
Only a few extreme observations are detected — only in XS3, XS9, and XS10 — confirming that the return distributions are generally stable but occasionally affected by rare, large shocks.
The relatively few outliers suggest that the winsorization applied later only has a marginal impact on the overall results.

Part 2

The second part of your task is to work out how sensitive each excess return is to all of the risk factors. For each return XS1 to XS10 in turn:

1) Run a regression of the excess return on a constant term, RMRF, SMB, HML, and UMD;
2) Interpret the regression coefficients, their t-ratios or p-values;
3) Store the slope coefficients form these regressions in an Excel file;
4) Comment on the variations in risk exposures across factors and how they might be interpreted;
5) Also interpret the R^2 from these regressions.

In this part of the assignment, we run a multiple linear regression to explain the excess return of each portfolio using the main risk factors from the Fama-French and Carhart models.

The model is:

$$R_i - R_f = \alpha_i + \beta_{i,M}(R_M - R_f) + \beta_{i,SMB}SMB + \beta_{i,HML}HML + \beta_{i,UMD}UMD + \varepsilon_i \qquad i = 1,\ldots,10$$

Here, $R_i - R_f$ represents the excess return of the portfolio i, that is the portfolio return minus the risk-free rate (calculated in the previous part)

The alpha (α) is a constant term capturing any average return that is not explained by the risk factors. The coefficients β_M, β_{SMB}, β_{HML}, β_{UMD} are called factor loadings or risk exposures, because they measure how sensitive the portfolio is to each source of systematic risk (for example, a high market beta means the portfolio moves more with the market, while a positive SMB beta means it behaves more like small-cap stocks. That's why these coefficients are called factor loadings or risk exposures: they quantify how exposed the portfolio is to each factor)

Interpretation of the factors:

- RMRF (Market factor): this represents the market risk premium, i.e., the return of the market portfolio minus the risk-free rate. A higher β_M means the portfolio moves more closely with the overall market, it carries more market risk exposure.

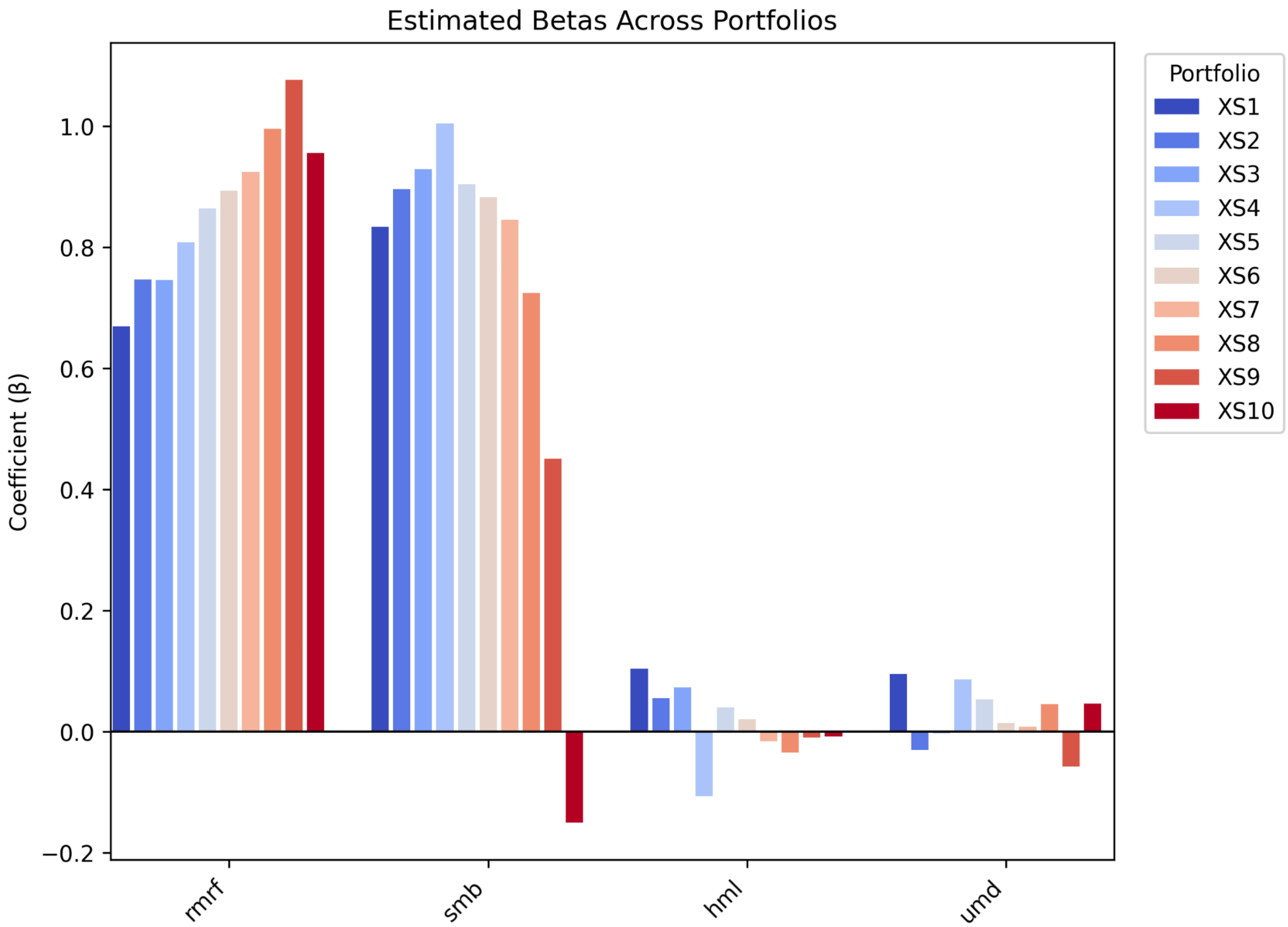
- SMB (Small Minus Big): The size factor captures the difference in returns between small-cap and large-cap stocks. A positive β_{SMB} indicates that the portfolio behaves more like small-cap stocks, while a negative value implies exposure to large-cap companies.

- HML (High Minus Low): This is the value factor: the return on value stocks (high book-to-market) minus growth stocks (low book-to-market). A positive β_{HML} suggests exposure to value stocks, while a negative one indicates exposure to growth stocks.

- UMD (Up Minus Down): The momentum factor measures the return of past winners minus past losers. A positive β_{UMD} means the portfolio has exposure to stocks with positive momentum — those that have recently performed well.

By running this regression for each portfolio (from XS1 to XS10), we can measure how its risk exposures vary across factors.
In other words, we see how much of each portfolio's performance can be explained by the market, size, value, and momentum effects, and how these exposures differ from one portfolio to another.

Estimated betas across portfolios
The figure below shows the estimated factor loadings (β coefficients) from the time–series regressions for all ten portfolios across the four factors.



The plot shows differences in risk exposures across the size-sorted portfolios. All portfolios load positively on the **market factor (RMRF)**, confirming that market risk is the main driver of excess returns.

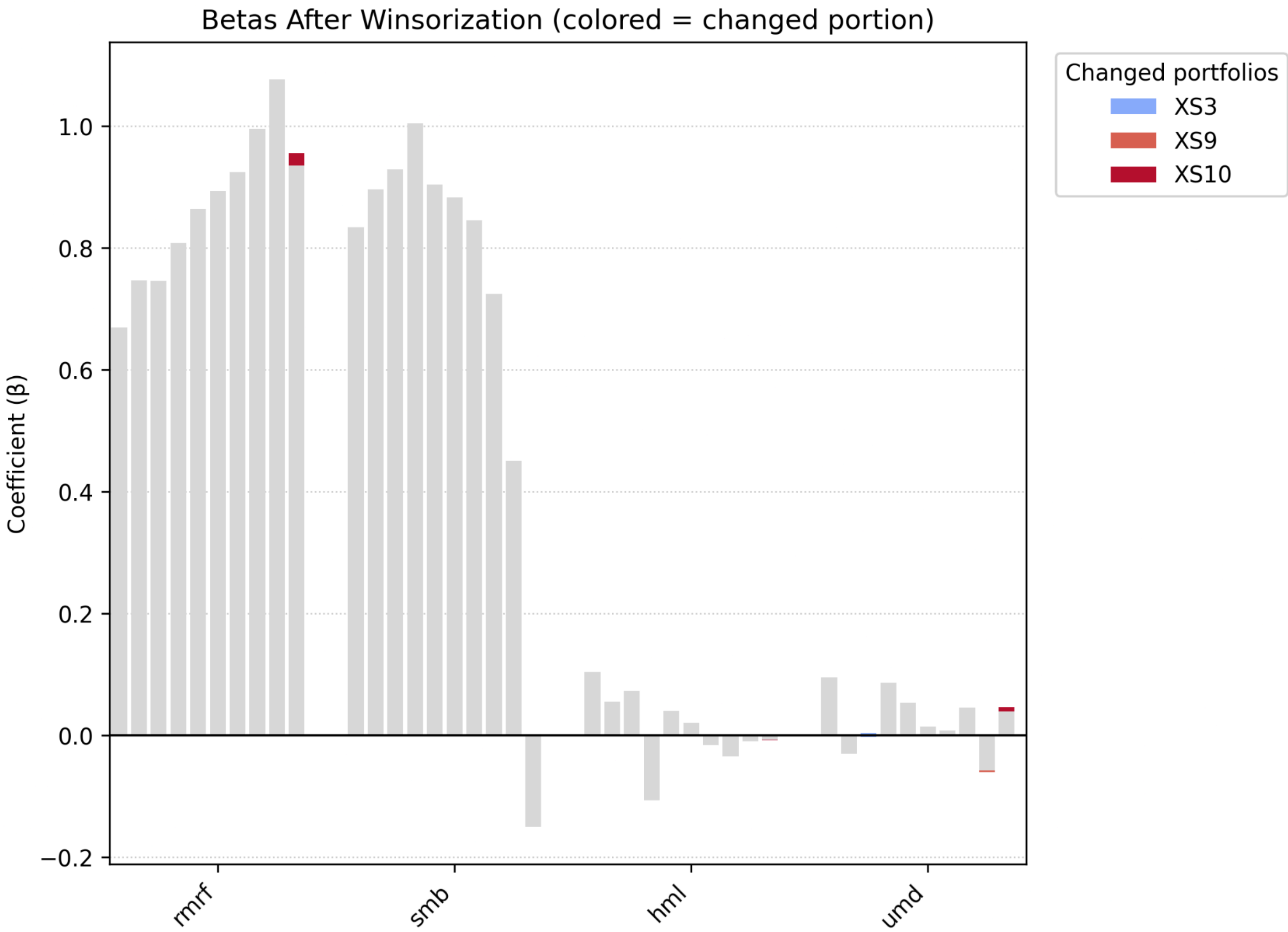
Interestingly, the market betas increase with portfolio size, meaning that large-cap portfolios are more correlated with the market than small-cap ones. Although this pattern appears counterintuitive from a theoretical perspective—since smaller firms are typically considered riskier—it is a common empirical outcome when using a market-cap–weighted index. In such a setup, the market factor naturally co-moves more with large-cap portfolios, while the higher volatility of small-cap portfolios is mainly idiosyncratic rather than driven

*by market movements.) (**not sure**)*
The **SMB (size)** decreases sharply from XS1 to XS10: small-cap portfolios exhibit higher sensitivity to the size premium, while large-cap portfolios have almost no exposure or even slightly negative loadings.
The **HML (value)** coefficients are close to zero across all portfolios, suggesting a neutral stance between value and growth stocks.
Similarly, the **UMD (momentum)** exposures are small and not systematic, indicating that the portfolios are not strongly affected by momentum effects.

Overall, the results suggest that **larger portfolios are more correlated with the overall market**, while **smaller portfolios are more exposed to the size premium**, consistent with the Fama–French framework.

Changes in Betas after winsorization:

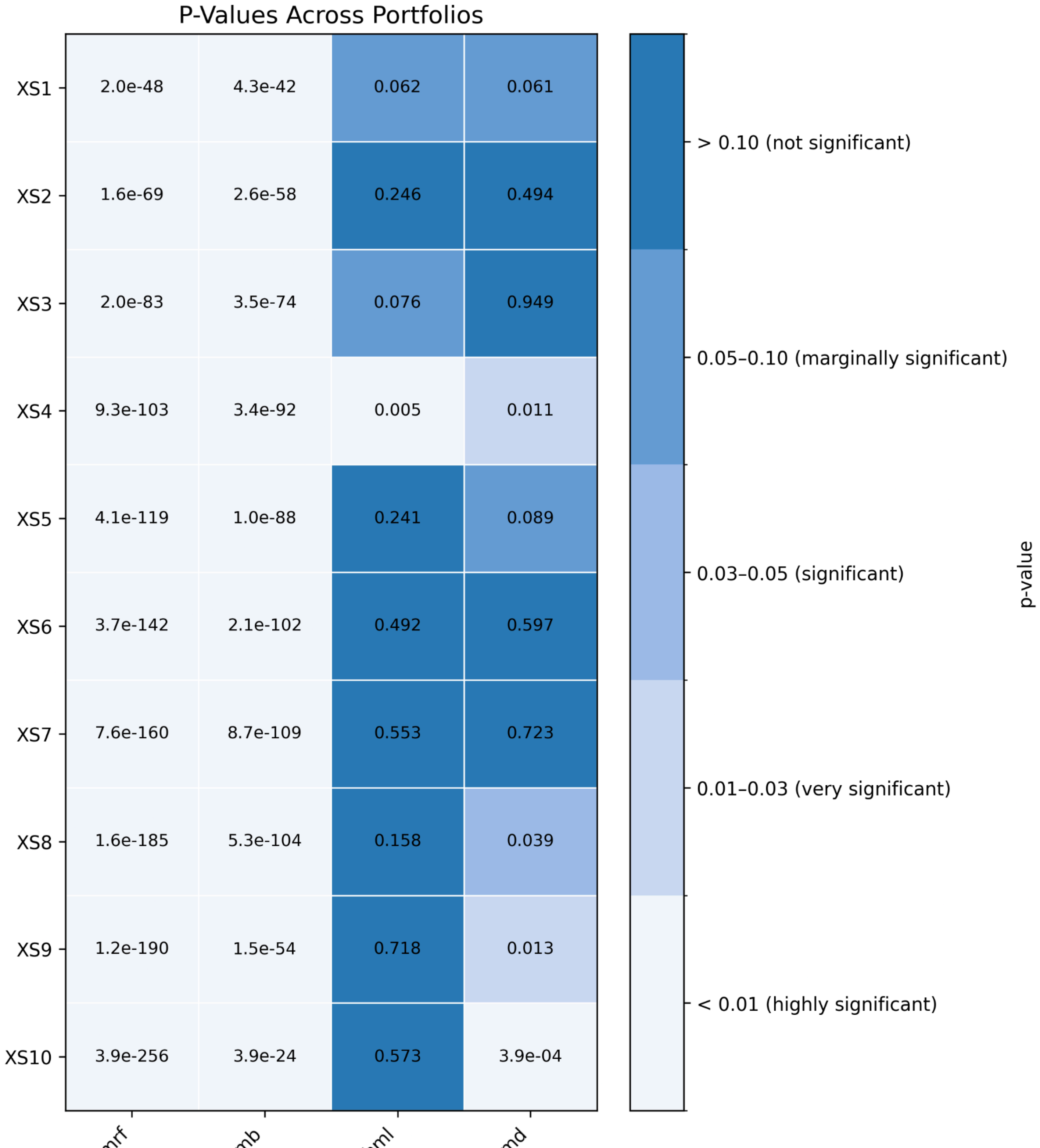
The figure below compares the estimated factor loadings before and after winsorization, with colored segments indicating the portion of each coefficient that changed.



As expected, the winsorization process — which trims extreme values in the data — has **negligible impact** on the estimated coefficients.

Only a few betas, mainly for portfolios XS3, XS9, and XS10 (the ones with outliers), show minimal adjustments, while the overall pattern and magnitude of exposures remain virtually unchanged.
This confirms that the results are **robust to outliers**, and that the original estimates were not driven by extreme observations.

P-Values across portfolios
The heatmap below reports the **p-values** from the time–series regressions for each portfolio and factor.
The p-value measures the **statistical significance** of a coefficient: it indicates the probability of observing a value at least as extreme as the estimated one, assuming that the true coefficient is zero.
In general, smaller p-values (typically below 0.05) suggest that the factor has a statistically significant effect on the portfolio's excess return.

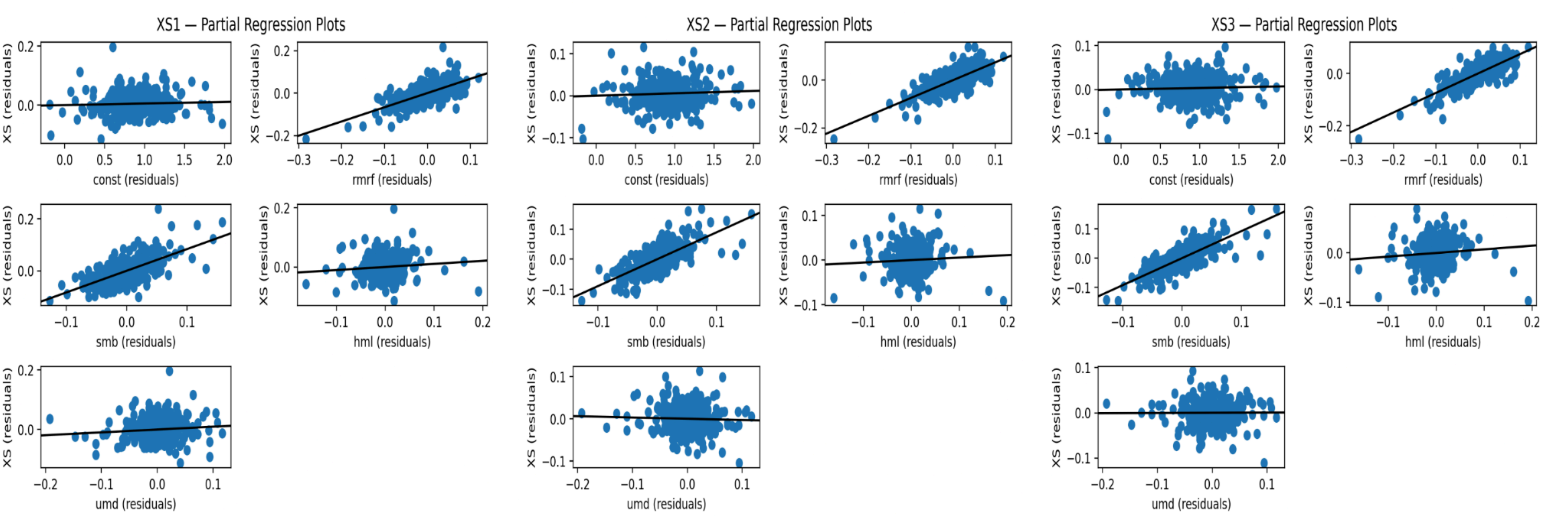


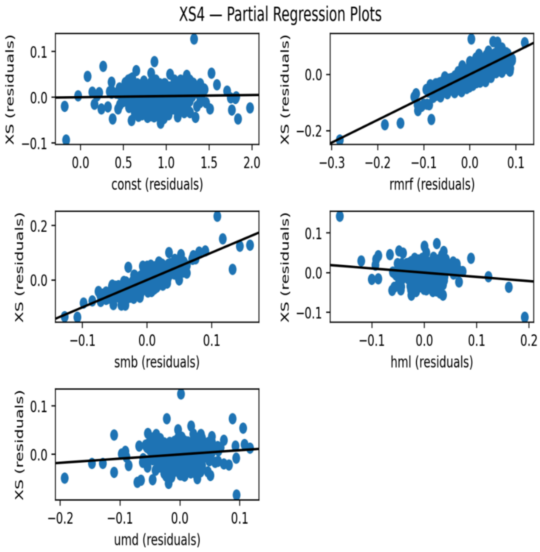
The figure shows that the **market factor (RMRF)** and the **size factor (SMB)** are highly significant for all portfolios, confirming that both factors play a key role in explaining excess returns across the size spectrum. In contrast, the **value factor (HML)** is mostly insignificant, except for a few marginal cases, indicating that value characteristics do not contribute much to portfolio performance.
The **momentum factor (UMD)** shows mixed results — it is significant for some of the larger portfolios (XS8–XS10) but not for the smaller ones — suggesting that momentum effects are more relevant among large-cap stocks.

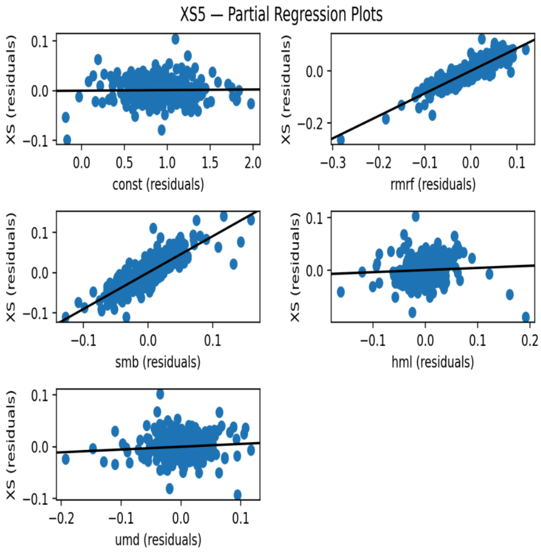
Overall, these results confirm that **market and size are the dominant drivers of returns**, while value and momentum play only a minor or sporadic role in this sample.

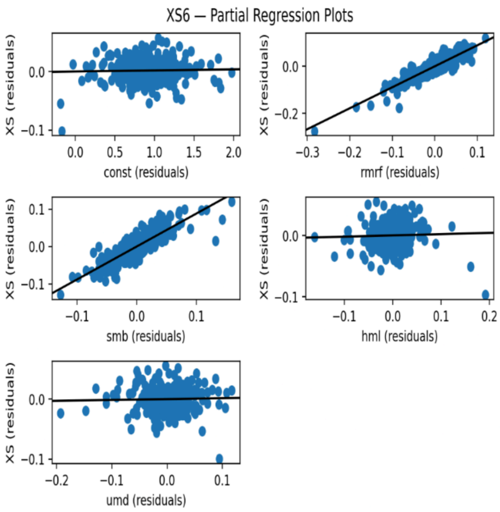
Partial Regression

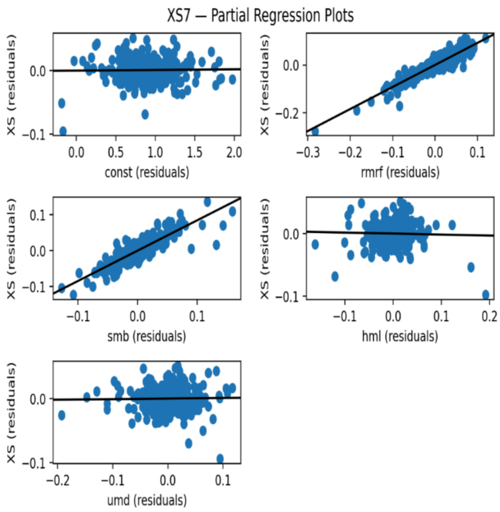
The figure below presents the partial regression plots for all ten portfolios and the four explanatory factors. Each plot isolates the relationship between the portfolio's excess return and a single factor, controlling for the influence of the others. In other words, a partial regression plot shows how much of the portfolio's return (after removing the effect of the other factors) can be explained by the residual component of one specific factor. This visualization helps to assess both the strength and direction of each factor's contribution to the regression.

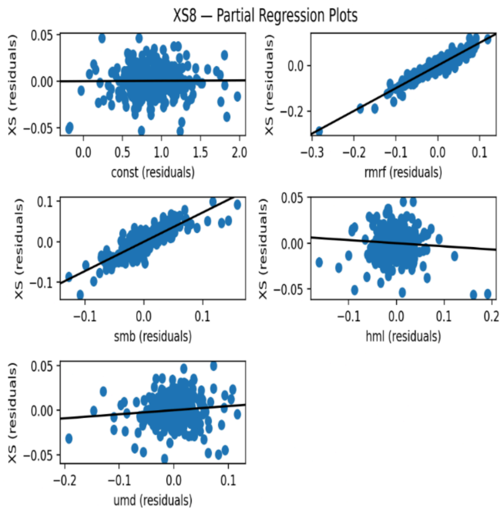


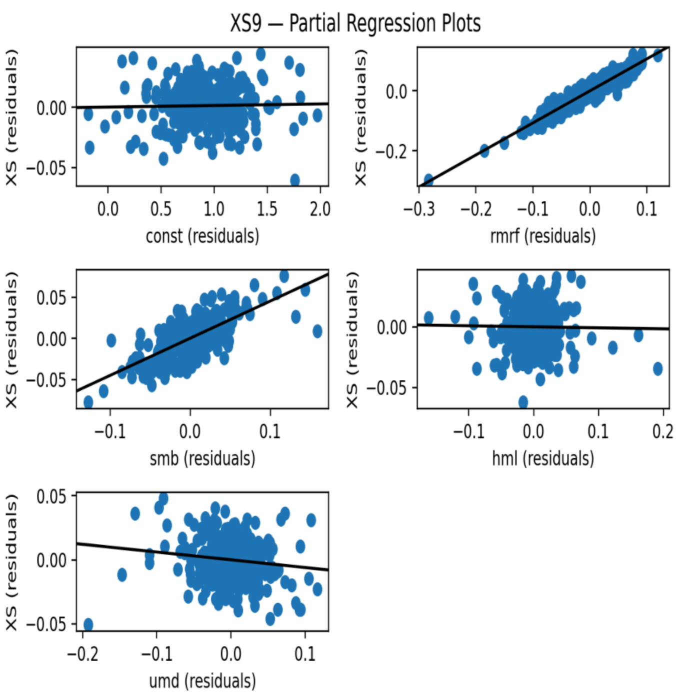

XS4 — Partial Regression Plots
XS (residuals)
const (residuals)
rmrf (residuals)
smb (residuals)
hml (residuals)
umd (residuals)

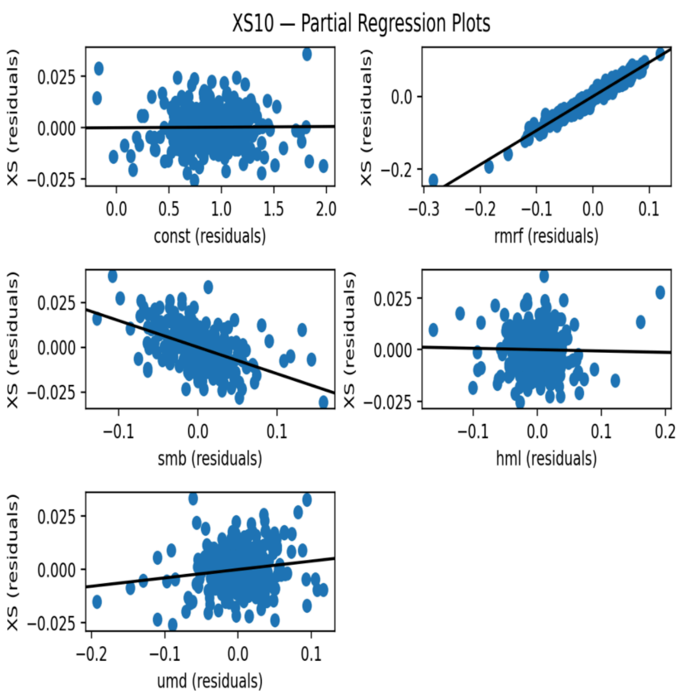

XS5 — Partial Regression Plots
XS (residuals)
const (residuals)
rmrf (residuals)
smb (residuals)
hml (residuals)
umd (residuals)


XS6 — Partial Regression Plots
XS (residuals)
const (residuals)
rmrf (residuals)
smb (residuals)
hml (residuals)
umd (residuals)


XS7 — Partial Regression Plots
XS (residuals)
const (residuals)
rmrf (residuals)
smb (residuals)
hml (residuals)
umd (residuals)


XS8 — Partial Regression Plots
XS (residuals)
const (residuals)
rmrf (residuals)
smb (residuals)
hml (residuals)
umd (residuals)


XS9 — Partial Regression Plots
XS (residuals)
const (residuals)
rmrf (residuals)
smb (residuals)
hml (residuals)
umd (residuals)


XS10 — Partial Regression Plots
XS (residuals)
const (residuals)
rmrf (residuals)
smb (residuals)
hml (residuals)
umd (residuals)

The plots confirm the strong and positive relationship between portfolio returns and the **market factor (RMRF)**. The dispersion of points around the regression line becomes narrower as we move from XS1 to XS10, meaning that **RMRF explains a larger share of the variation** for larger portfolios — consistent with their stronger exposure to overall market movements.
The **size factor (SMB)** shows a clear positive relationship for the small-cap portfolios, which progressively weakens for larger ones.
In contrast, the **value (HML)** and **momentum (UMD)** factors display flatter and more scattered relationships, indicating that their contribution to explaining returns is minor or not statistically significant.

The intercept (α) is not statistically significant for any portfolio, which suggests that the four-factor model adequately explains the average excess returns. In other words, there is no evidence of abnormal performance once the common risk factors are taken into account.

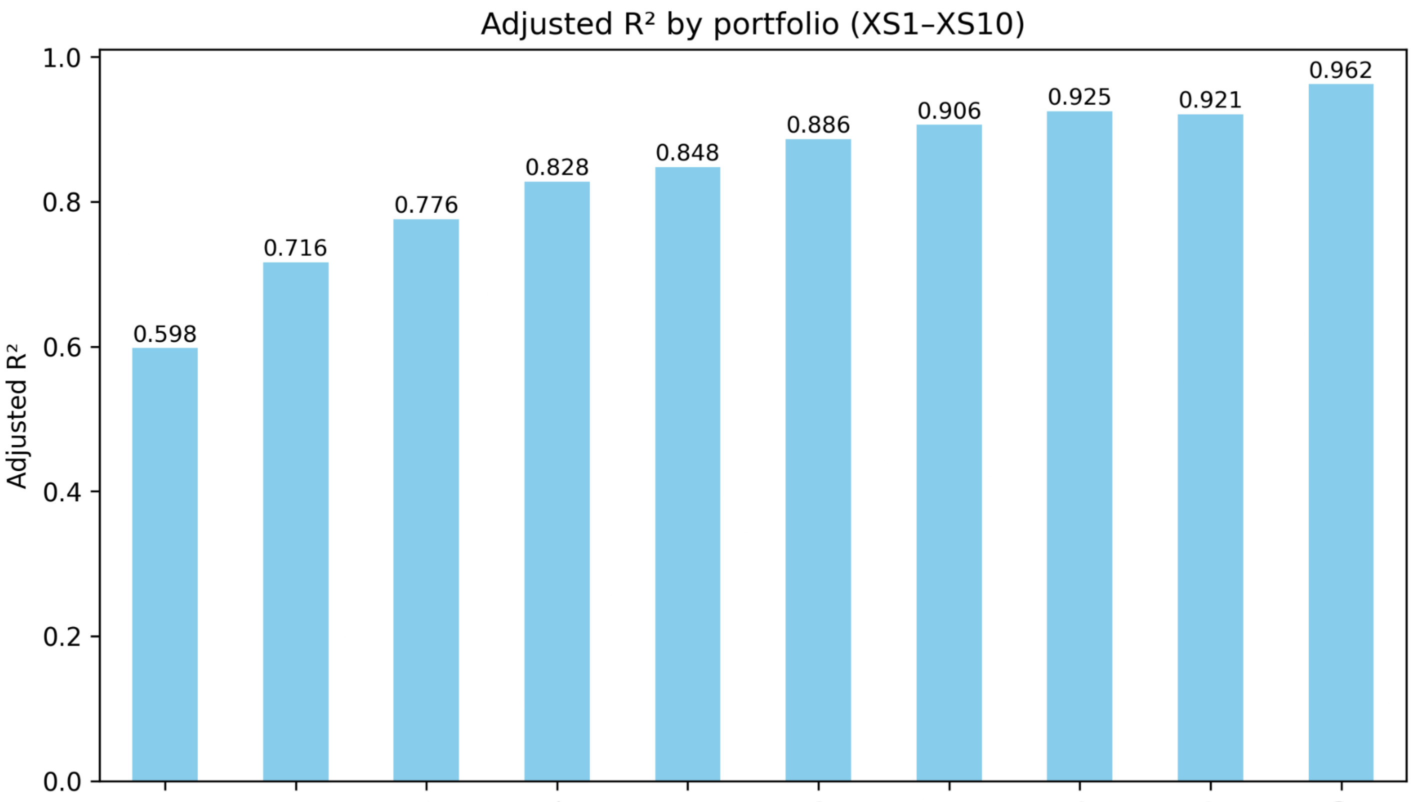
Overall, the partial regression plots confirm that **market exposure dominates in large-cap portfolios**, while the **size effect** remains the key driver for small-cap ones.

Adjusted R^2 by portfolios

The figure below shows the Adjusted R^2 for each of the ten size-sorted portfolios.

The Adjusted R^2 measures how well the regression model explains the variation in excess returns, while also accounting for the number of explanatory variables used.

Unlike the regular R^2, which can only increase when new variables are added, the Adjusted R^2 penalizes the inclusion of factors that do not improve the model's explanatory power — making it a more reliable measure for model comparison.



The Adjusted R^2 values increase steadily from XS1 (0.60) to XS10 (0.96), indicating that the **model fits the data better for larger portfolios**.
This means that the four Fama–French–Carhart factors — market, size, value, and momentum — explain a greater proportion of the excess return variability for large-cap portfolios than for small-cap ones.
The lower R^2 values for small-cap portfolios suggest the presence of **additional idiosyncratic or non-systematic risks** that are not captured by the model.

Overall, the high Adjusted R^2 values across all portfolios confirm that the model has strong explanatory power, especially for larger firms whose returns are more closely linked to common market factors.

This pattern is consistent with the idea that small firms are more heterogeneous and subject to firm-specific shocks, whereas large firms move more in line with systematic market dynamics.

*The exceptionally high Adjusted R^2 for portfolio XS10 can be partly explained by the fact that the market factor (RMRF) is market-cap weighted. Since large-cap stocks dominate the market portfolio, the returns of XS10 are naturally very similar to those of the market itself, leading to a nearly perfect fit (**not sure, have to ask to prof. if the market portfolio is market-cap weighted or equally weighted**).*

Part 3

The final part of our job. Do risk exposures explain the way in which mean returns vary across portfolios?

1. Run a series of cross section regressions. The y-variable is the set of 10 mean returns on the ten size-based portfolios, and the x-variables are a constant and two sets of risk exposures from your time series regressions; those on RMRF and one other. In turn, the other risk factors will be SMB, HML, and UMD. Thus, in each regression, there are 10 observations and two explanatory variables plus a constant term;
2. Interpret your results;

In the last part of our job, we run a series of **cross-sectional regressions** in order to examine whether the risk exposures estimated in the previous time-series analysis are able to explain the differences in the **average excess returns** across the ten size-sorted portfolios.
The estimated betas from the time-series regressions are treated as explanatory variables, while the dependent variable is the set of mean excess returns for the ten portfolios.

Specifically, we estimate three separate cross-sectional models, each including the market factor RMRF and one additional factor at a time (SMB, HML, UMD).
The model can be expressed as:

$$\overline{r}_i = \lambda_{const} + \lambda_M \beta_{i,M} + \lambda_X \beta_{i,X} + \varepsilon_i \qquad i = 1,\dots, 10$$

where:

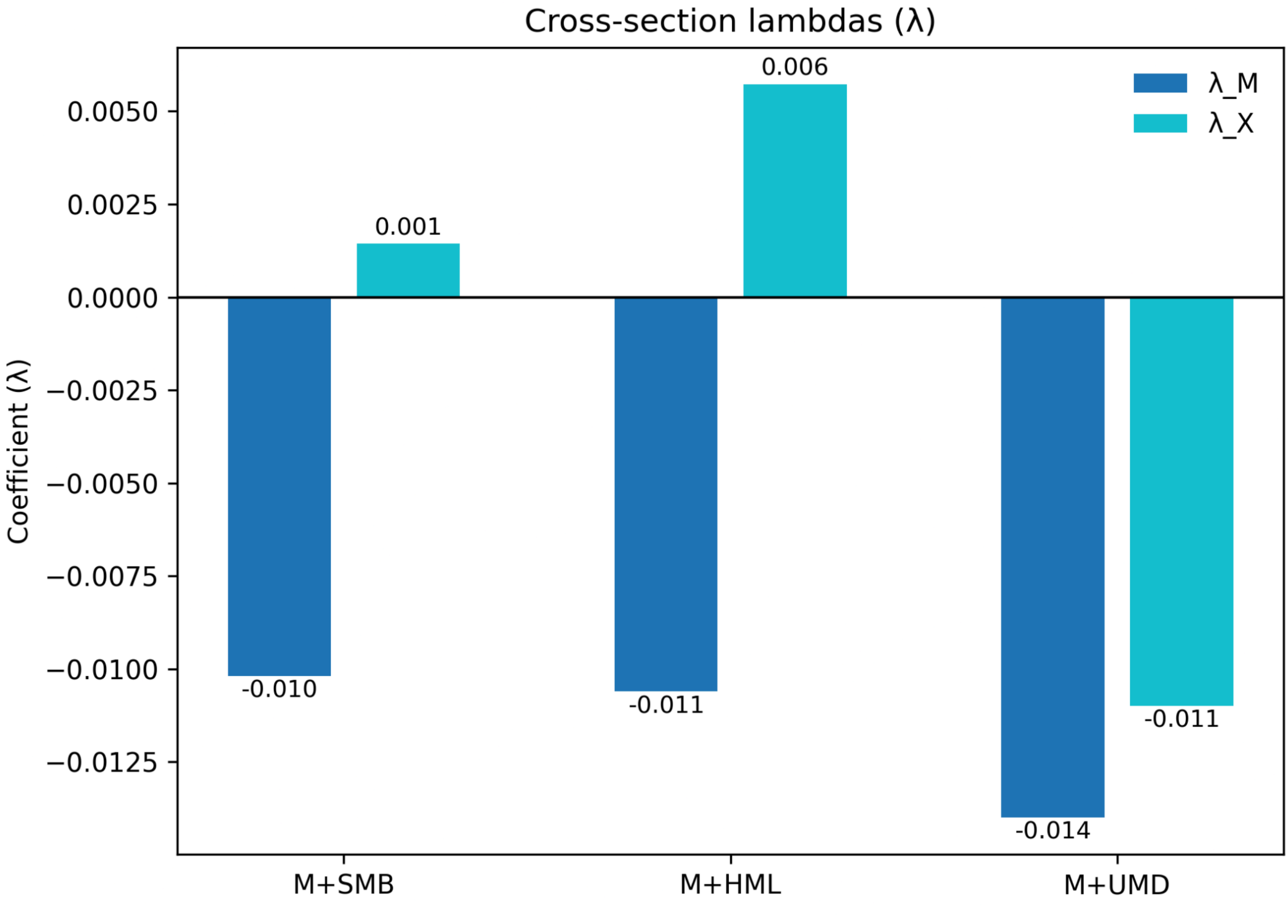
- $\overline{r}_i$ is the **mean excess return** of portfolio i,
- $\beta_{i,M}$ and $\beta_{i,X}$ are the **risk factors** (estimated in the previous time-series step),
- λ_{const} is the intercept,
- λ_M and λ_X represent the **risk premium** associated with the market factor and with factor $X \in \{SMB, HML, UMD\}$, showing how the average returns vary with each factor,
- and ε_i is the pricing error for portfolio i.

Each regression therefore contains **10 observations** and **3 parameters** (the intercept and two factor coefficients).
Comparing the estimates of λ_M and λ_X across the three specifications allows us to assess which factors are priced in the cross-section of returns and how well the models explain the differences among the portfolios.

Interpretation of the estimated λ coefficients

The figure below reports the estimated coefficients λ_M and λ_X from the three cross-sectional regressions, where X represents in turn the SMB, HML, and UMD factors.

Each λ measures how the average excess returns of the ten size-based portfolios vary with respect to the corresponding factor exposures obtained in the previous time-series step.



Across all specifications, the **market coefficient** λ_M is **negative**, indicating that portfolios with higher market betas tend to exhibit **lower average excess returns** in the cross-section.

This result is consistent with the pattern observed in size-sorted portfolios: small-cap portfolios generally show **higher average excess returns** but **lower market betas**, while large-cap portfolios are **less risky** and yield **lower excess returns**.

As shown in Part 2, the estimated market betas increase with portfolio size, meaning that large-cap portfolios are more correlated with the market. Consequently, the cross-sectional relationship between market beta and mean return becomes negative, even though this does not necessarily contradict the theoretical positive risk–return trade-off of the CAPM.

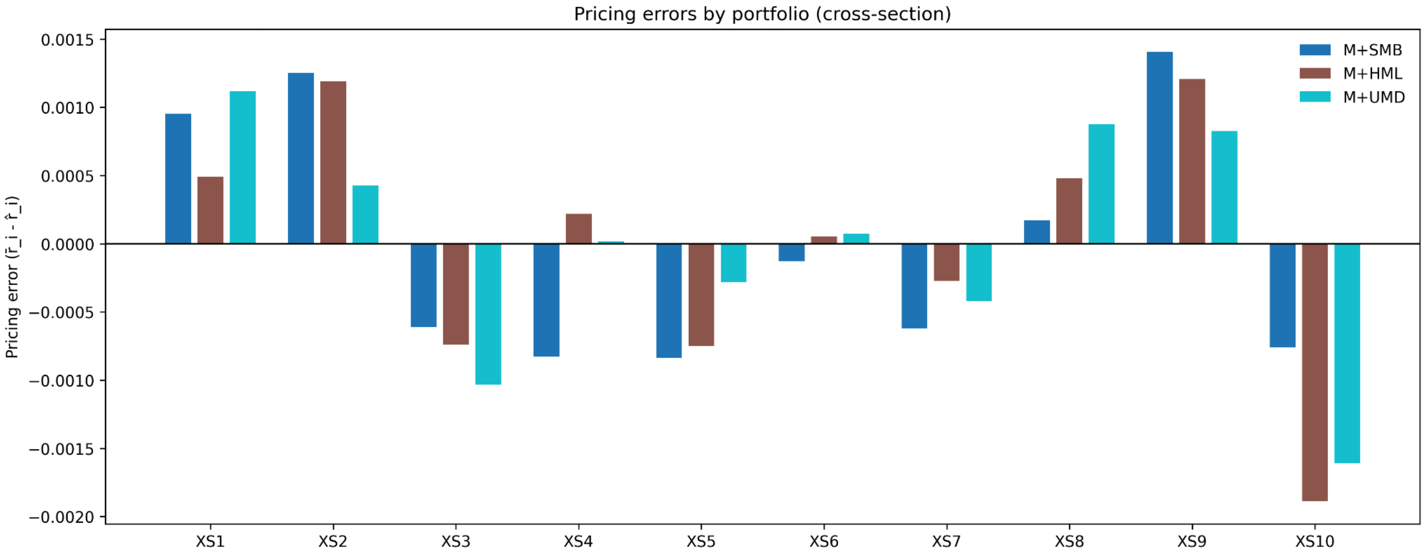
As for the additional factors, the results are mixed:

- In the **M+SMB** regression, λ_{SMB} is small and positive, indicating that the size factor contributes only marginally once the market exposure is controlled for.
- In the **M+HML** model, λ_{HML} is positive and somewhat larger, suggesting that portfolios with greater exposure to the value factor tend to earn slightly higher average returns.
- In the **M+UMD** specification, both coefficients are negative, implying that neither the market nor the momentum factor explains much of the cross-sectional variation in average excess returns.

Overall, the estimated λ coefficients are small in magnitude and vary in sign across factors, indicating that these simple two-factor specifications provide **limited explanatory power** when applied to only ten size-based portfolios.

Pricing errors across portfolios

The chart below reports the pricing errors from the three cross-sectional regressions, computed as the difference between the observed and fitted mean excess returns $\overline{r}_i - \widehat{r}_i$



These errors represent the portion of average returns that is not captured by the corresponding factor specification.

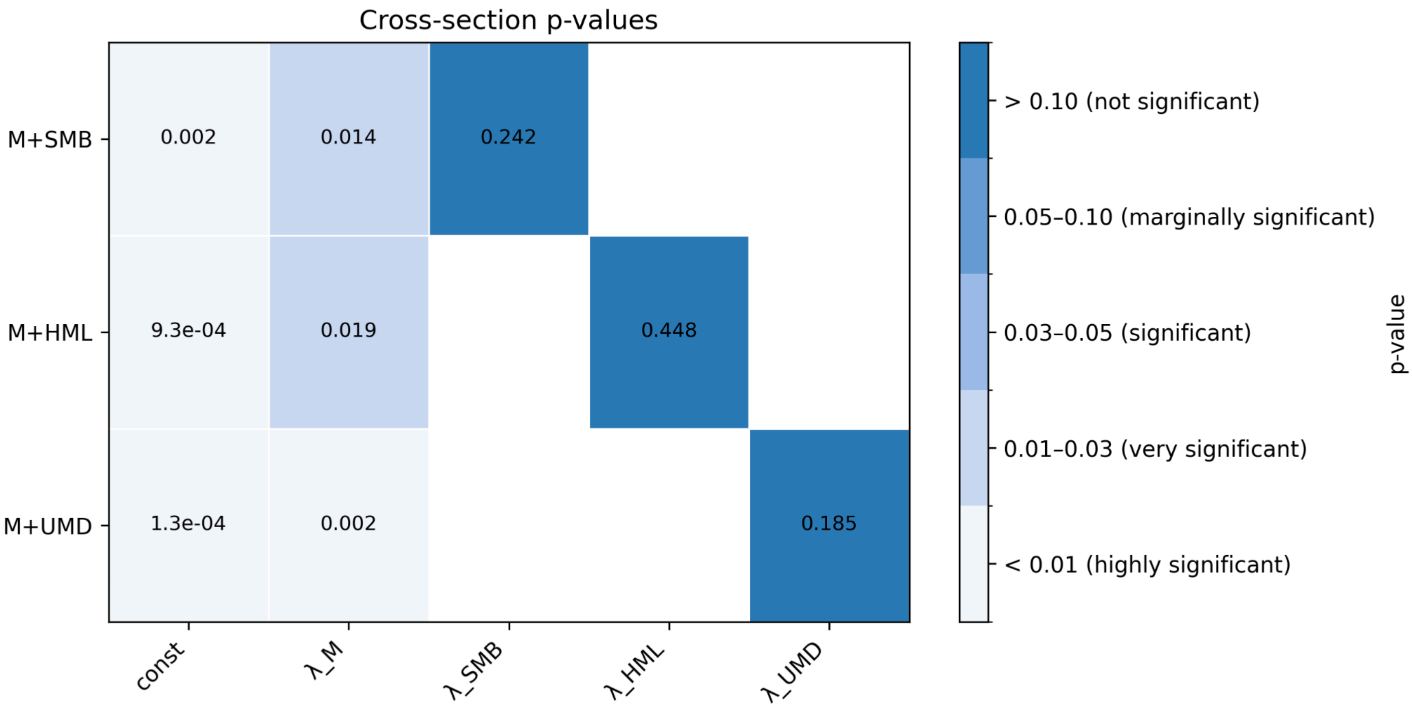
Overall, the pricing errors fluctuate around zero but show clear variation across portfolios.
The largest deviations appear at the extremes of the size distribution — particularly for the smallest and largest portfolios — suggesting that the two-factor models struggle to fully explain the average returns of very small or very large firms.
Among the three specifications, none provides a uniformly better fit across all portfolios, although the **M+HML** model tends to produce slightly larger errors for the largest portfolios, while **M+SMB** and **M+UMD** display similar patterns.

The alternating positive and negative residuals indicate that the fitted values systematically under- and over-predict returns for different segments of the size spectrum, pointing to potential **omitted factors** or **nonlinearities** that the linear cross-sectional regressions cannot capture.
This evidence confirms that, even though factor exposures explain part of the cross-sectional variation in average excess returns, a substantial fraction remains unexplained by these simple models.

Significance of the cross-sectional coefficients

The heatmap below reports the p-values associated with the estimated coefficients from the three cross-sectional regressions.

Darker shades correspond to higher p-values and therefore lower statistical significance.



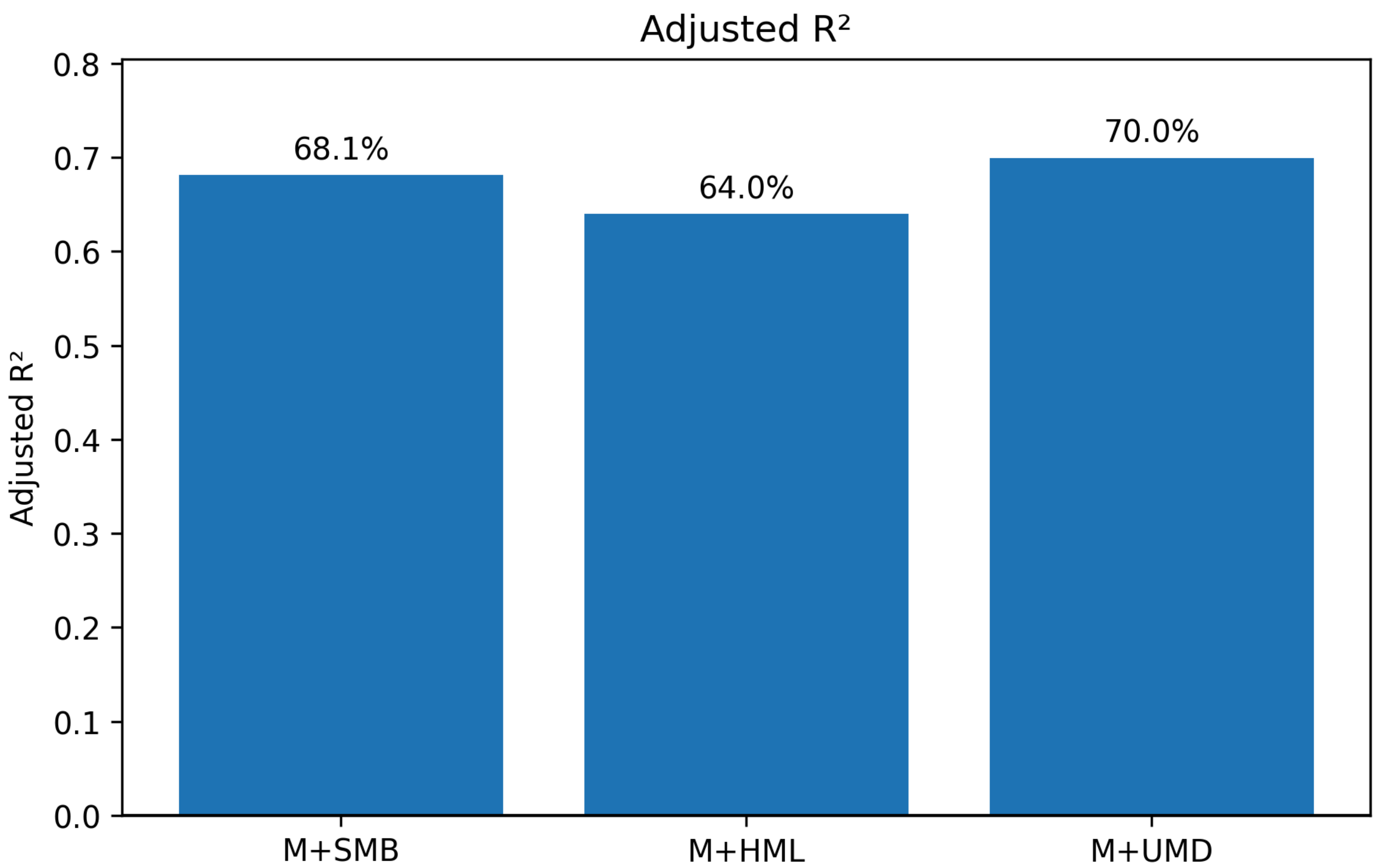
The results show that the intercept and the market coefficient λ_M are statistically significant in all specifications, with p-values well below the 5% threshold.
This confirms that the market factor remains the dominant source of cross-sectional variation, even though its estimated λ is negative.

In contrast, the additional factors are not statistically significant.
The SMB, HML, and UMD coefficients all display p-values above 0.10, indicating that none of them provides a meaningful improvement in explaining average excess returns across the ten portfolios.
The lack of significance suggests that, within this limited cross-section, the size, value, and momentum effects are not sufficiently strong to emerge once the market exposure is accounted for.

Overall, the pattern of p-values reinforces the previous interpretation: the market factor drives most of the variation, while the other factors add little explanatory power when applied to size-sorted portfolios over this sample period.

Explanatory power of the cross-sectional models

The figure below compares the adjusted R^2 values obtained from the three cross-sectional regressions. These statistics measure how much of the variation in average excess returns across the ten portfolios is explained by each factor specification, after adjusting for the number of regressors.



All models achieve relatively high adjusted R^2 values, ranging between 64% and 70%, indicating that the market factor and one additional exposure can jointly capture a substantial portion of the cross-sectional differences in mean returns.
Among the three, the **M+UMD** regression delivers the highest explanatory power, followed closely by **M+SMB**. The **M+HML** model performs slightly worse, suggesting that the value factor contributes less to explaining the cross-section of average returns in this sample.

However, despite these moderately high R^2 values, the results from the pricing errors and p-value analysis show that the coefficients are not always statistically significant and that large pricing deviations remain for some portfolios.
Therefore, while the models provide a reasonable fit in aggregate, they still fail to fully account for the cross-sectional variation in mean excess returns.