R&D PORTFOLIO EXPLORATION: EXECUTIVE MEMORANDUM

Medha Nalamada: mrn789
DATE: Dec 9 2024
SUBJECT: R&D Portfolio Testing

Introduction & Objective

This report examines the performance of firms with and without reported R&D expenditures through the construction and evaluation of value-weighted and equal-weighted portfolios. Using data spanning from 1980 to 2022, the portfolios' alphas and betas relative to the market are found and informed whether this factor can explain the variation in returns that other factors like small-minus-big, or high-minus-low might have missed. By incorporating WRDS datasets such as Compustat fundamentals, CRSP stock market returns, and linking tables, this exploration provides a comprehensive evaluation of whether R&D expenditure influences portfolio performance, and understanding how it shapes returns can lead to many different potential investment strategies for U.S. firms in the future.

Methodology

Data Acquisition and Cleaning

The Wharton Research Data Services (WRDS) database is a widely utilized data platform designed to support research in a multitude of disciplines. For this research, data was acquired from two WRDS datasets: Compustat/CRSP linked fundamentals annual data and CRSP (V2) stock monthly returns data. Compustat provided firm-level financial data, including R&D expenditures, while CRSP offered market return and capitalization details. The two data sets were merged on the year and a unique identifier (Year and PERMNO).

The Compustat dataset, which contains firm-level financial and accounting data, was filtered to include only U.S.-domiciled firms using USD currency. This restriction ensures that all firms in the dataset operate under similar economic and regulatory environments, avoiding inconsistencies arising from currency fluctuations or international accounting standards. Furthermore, firms in the financial (SIC 6000–6999) and pharmaceutical (SIC 2834) sectors were excluded. Financial firms are typically regulated differently and have unique capital structures, making them less comparable to

firms in other industries. Similarly, pharmaceutical firms often exhibit extreme R&D spending patterns that could skew the analysis. By excluding these sectors, the dataset better represents the general landscape. Additionally, to align fiscal year data with calendar years, the fiscal year-end was shifted forward by 90 days. This adjustment ensures that firm-level accounting data corresponds to the correct calendar year's market performance. For example, a firm with a fiscal year ending on September 30th, 2022, would have its financial data aligned with calendar year 2023, reflecting the period when investors likely react to these financial disclosures. A "next year" variable was created to operationalize this alignment, advancing the fiscal year by one and ensuring that R&D expenditures and other fundamentals from Compustat corresponded to the subsequent year's market returns in CRSP. For the 'xrd' (R&D expense variable), rows with no revenue and no xrd expense were removed. If a firm did not have revenue included, they might not have reported their financials, and assuming xrd as 0 (counting it for the non-R&D group) if financials were missing could serve as a misleading assumption. Hence, rows with revenue and 'xrd' missing were dropped, and now rows with xrd > 0 were counted for the R&D portfolio.

The CRSP dataset, which provides detailed market data, underwent a similar filtering process. In addition to the SIC Code filtering, observations with invalid or missing returns were removed to avoid inaccuracies in the computation of portfolio returns. Extremely negative returns (e.g., less than -100%) were also excluded. To mitigate lookahead bias, market capitalization values were shifted by one month. This ensures that portfolio weights are based on information available at the time of portfolio construction, aligning the methodology with real-world investment practices.

The market factor data was loaded from the Kenneth R. French Data Library. This included the risk-free rate and excess market return factor for all months and was filtered accordingly for the date range.

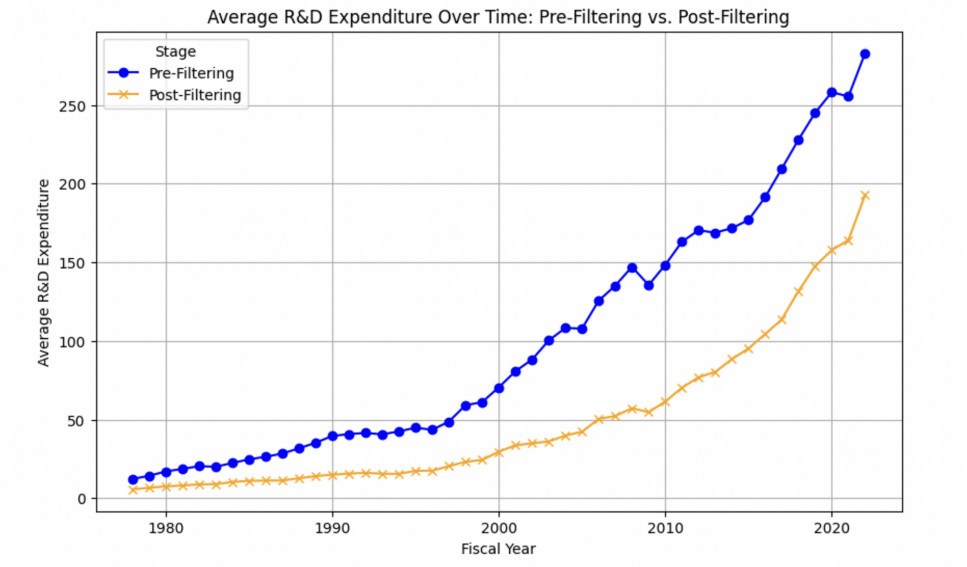
Merging and Validation

The cleaned datasets were merged together based on the available year and the unique identifier PERMNO. The merging process ensured that only valid link dates were retained, and the Compustat and CRSP data were combined. The resulting dataset was also merged to include the factor data and filtered down to contain firm-level data on returns, market capitalization, and R&D expenditures across the specified time period 1980-2022.

EDA & Characteristics

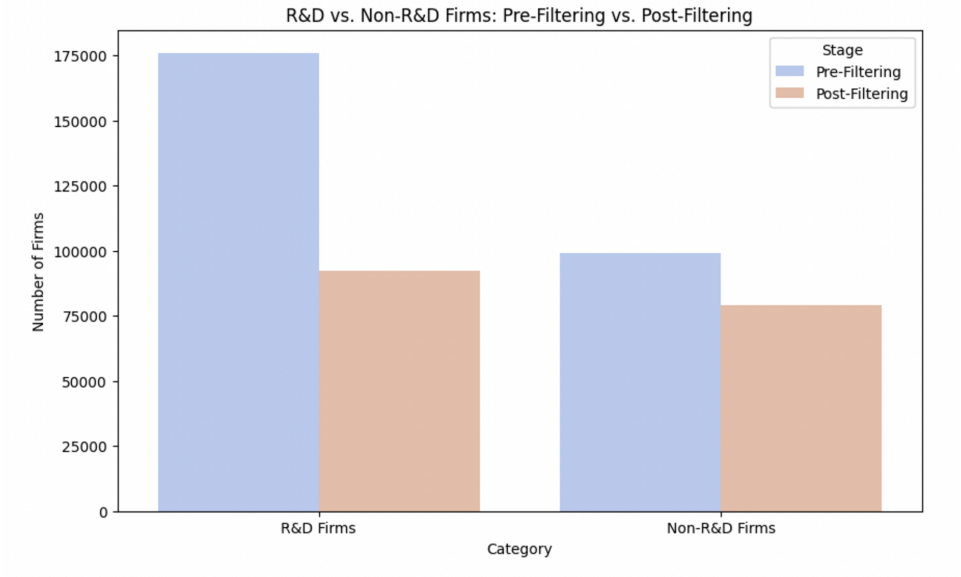
The exploratory data analysis (EDA) provided critical insights into the dataset's structure, characteristics, and transformations during cleaning and merging. Pre-cleaning, the dataset contained a substantial number of firms, but strict filtering criteria (e.g., exclusion of financials, pharmaceuticals, non-USD firms, and extreme values) led to a significant reduction in observations.

Figure 1: Average R&D Expenditure over time



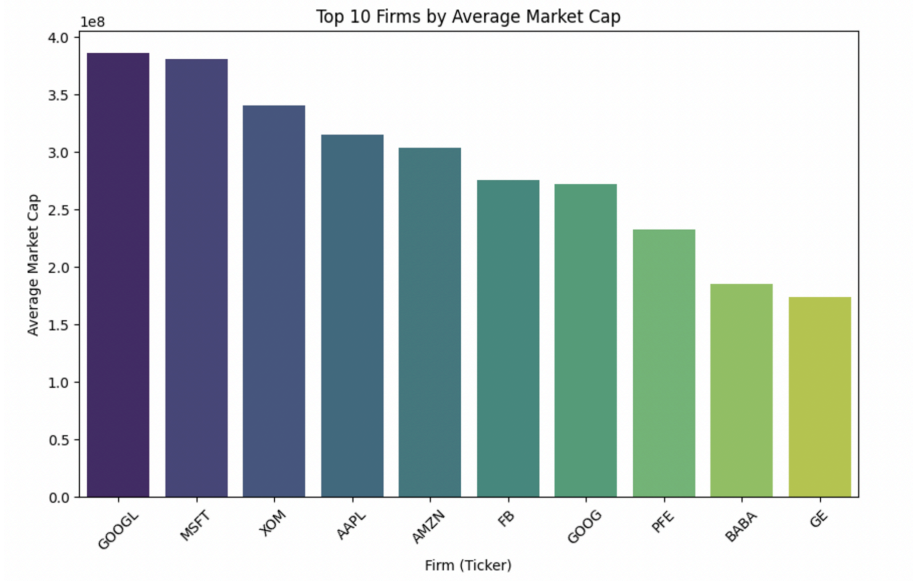
The chart illustrates the average R&D expenditure over time, comparing pre-filtering and post-filtering datasets. Both datasets show a consistent upward trend in R&D spending from the 1980s to the 2020s, reflecting a growing emphasis on innovation. However, the post-filtering values are consistently lower, indicating that filtering removed firms with disproportionately high R&D expenditures, such as financials, pharmaceuticals, and non-USD entities. Post-filtering data shows a more gradual and linear growth pattern, suggesting a more representative subset of firms. This subset selection significantly reduced the amount of firms that had R&D vs. didn't have R&D, but the filtering ensured both groups were balanced, making the insights from the regression more reliable, as shown in Figure 2.

Figure 2: R&D vs Non-R&D Firms



The unique firm data (Exhibit #1) across time also varied, with the highest number of unique firms available occurring between 1990 and 2000. The number is lower in more recent years, but still sufficient to make reliable insights. As of 2023, the largest firms by market capitalization are presented in Figure 3.

Figure 3: Top 10 Firms by Market Cap (Latest)



The graph highlights the top 10 firms by average market capitalization, showcasing the dominance of large-cap firms such as Alphabet (GOOGL, GOOG), Microsoft (MSFT), and Apple (AAPL). These firms hold significantly higher market values compared to others in the dataset, reiterating the disproportionate influence of a few large-cap firms on aggregate market trends. This emphasizes the need to distinguish between equal-weighted and value-weighted portfolio returns. Equal-weighted portfolios ensure each firm contributes equally, highlighting the performance of smaller firms, while value-weighted portfolios emphasize the influence of larger firms, offering a more market-representative perspective. The subsequent section will detail the methodologies used to create these portfolios.

Portfolio Construction

Two types of portfolios were constructed: equal-weighted and value-weighted. Firms were categorized into two groups based on their R&D expenditure (XRD): those with R&D (XRD>0) and those without (XRD=0). Equal-weighted portfolio returns were calculated as the average return of all firms in a group on a given date, while value-weighted portfolio returns were calculated by weighting each firm's return by its market capitalization. This distinction captures two different dimensions of portfolio performance: equal-weighted portfolios emphasize the performance of smaller firms by giving all firms the same weight, while value-weighted portfolios reflect the influence of larger, more established firms that dominate market capitalization. Market factors, including the risk-free rate (RF) and market excess return (RMRF), were incorporated into the dataset to calculate excess returns for each portfolio. Excess returns were derived as the difference between portfolio returns and the risk-free rate.

Analysis

CAPM Framework

To evaluate the performance of the constructed portfolios, a CAPM-style regression was employed. The regression model is as follows: $R_p = \alpha + \beta(RMRF_t)$

R_p represents the portfolio's excess return, $RMRF_t$ is the market excess return, α is the intercept (indicating the portfolio's alpha), and β is the slope (indicating market sensitivity). Separate regressions were conducted for each of the four portfolios: equal-weighted and value-weighted portfolios for both R&D and non-R&D firms. Key

regression statistics, including alpha, beta, R-squared, and significance levels, were extracted.

Figure 4: Regression Results

	Portfolio	Alpha	Alpha SE	Alpha P-Value	Beta	Beta SE	Beta P-Value	R-squared
0	Equal-Weighted, No R&D	0.000987	0.001379	0.474	1.062982	0.029986	p < 0.001	0.709713
1	Value-Weighted, No R&D	0.000228	0.000598	0.703	0.907673	0.012997	p < 0.001	0.904662
2	Equal-Weighted, R&D	0.001399	0.001771	0.430	1.260622	0.038522	p < 0.001	0.675691
3	Value-Weighted, R&D	0.000244	0.000694	0.726	1.082118	0.015101	p < 0.001	0.909005

Equal-Weighted Portfolio, No R&D:

The alpha for the equal-weighted portfolio of firms without R&D expenditure is 0.000528, with a standard error of 0.001375 and a p-value of 0.701, indicating that the alpha is not statistically significant. This suggests that there is no evidence that the portfolio's returns deviate from the market after adjusting for risk. The beta, measuring the portfolio's sensitivity to the market, is 1.073307, with a standard error of 0.029902 and a highly significant p-value of <0.001. This indicates that the portfolio has a positive and strong correlation with market returns. The R-squared value of 0.714820 suggests that a substantial portion of the portfolio's return variability is explained by market movements.

Value-Weighted Portfolio, No R&D:

For the value-weighted portfolio of non-R&D firms, the alpha is 0.000252, with a standard error of 0.000573 and a p-value of 0.661, again showing no statistical significance. This implies that the portfolio's performance is consistent with market expectations. The beta is 0.920001, with a standard error of 0.012463 and a significant p-value of <0.001, indicating a slightly lower sensitivity to market returns compared to its equal-weighted counterpart.

Equal-Weighted Portfolio, R&D:

The equal-weighted portfolio of firms with R&D expenditure has an alpha of 0.000411, with a standard error of 0.001792 and a p-value of 0.818, indicating a lack of statistical significance. The beta for this portfolio is 1.273987, with a standard error of 0.038966 and a highly significant p-value of <0.001, showing stronger market sensitivity than any of the other portfolios.

Value-Weighted Portfolio, R&D:

The value-weighted portfolio of firms with R&D expenditure reports an alpha of -0.000307, with a standard error of 0.000668 and a p-value of 0.646, indicating no statistical significance. The beta is 1.076874, with a standard error of 0.014537 and a highly significant p-value of <0.001, suggesting a strong positive relationship with market returns. The R-squared value of 0.914359 shows that market movements explain a large proportion of the portfolio's return variability.

Key Takeaways:

Alphas: None of the portfolios exhibit statistically significant alphas, suggesting no evidence of consistent outperformance or underperformance relative to the market after adjusting for risk.

Betas: All portfolios have significant betas, indicating strong correlations with market movements. R&D portfolios, particularly equal-weighted ones, show higher sensitivity to the market (beta > 1), suggesting that these portfolios are slightly more volatile compared to the market.

R-squared Values: Value-weighted portfolios consistently have higher R-squared values than equal-weighted ones, highlighting that larger-cap firms' returns are more strongly tied to market movements.

Team-Based Modifications and Future Considerations

After consolidating with the teams across class, the following revisions were made:

- Use of CRSP Version 2 Data Set. This new data set included a variable for market capitalization (MthCap) that was used instead of manually calculating price * shares outstanding, along with new variable names.
- Dropping Compustat rows with missing revenue. This policy was adjusted to ensure that setting xrd to 0 (if data was missing) was not overestimating firms with no R&D. Instead, if the firms were missing revenue, we concluded that the firm must have been missing financial data and hence excluded them from the analysis.
- Adjusting the Fama-French factors. In my initial exploration, the factor data was not adjusted to match the units of the Compustat/CRSP data. This was initially skewing my alpha values to be extremely high, and correcting them showed more interpretable results.

- Reconstruction of value-weighted portfolios. Initially, the value-weighted returns were being added as a new column, with a new entry for each row and were not being grouped by PERMNO. After adjusting this, the results were as expected.

For future analyses, several refinements can be explored to enhance the analysis. Incorporating additional factors beyond the market factor (RMRF), such as size (SMB) and value (HML) from the Fama-French three-factor model, could provide deeper insights into the drivers of portfolio performance. Additionally, testing alternative weighting schemes, such as equal-weighted portfolios with constraints or dynamically adjusted weights, may offer a better understanding of portfolio behavior. Finally, this exploration was limited to firms in the U.S. and had strict filtering policies; expanding the dataset to include international firms could provide a comparative perspective on R&D vs. Non-R&D performance across different economic environments.

Conclusion

The analysis highlights the distinct performance characteristics of portfolios comprising R&D and Non-R&D firms. While none of the portfolios exhibited statistically significant alphas, suggesting no consistent outperformance relative to the market after adjusting for risk, the beta values were significant across all portfolios, demonstrating strong correlations with market returns. R&D portfolios, particularly the equal-weighted ones, displayed higher sensitivity to market movements, as reflected by their beta values exceeding one. This suggests that smaller R&D-intensive firms contribute greater volatility and potentially higher returns. Conversely, value-weighted portfolios, dominated by larger firms, showed more stable performance with higher R-squared values, indicating stronger market dependence. These results emphasize the importance of R&D investments in influencing market behavior while highlighting the need for further exploration into additional factors that could impact portfolio returns. With the expansion of AI into the financial industry and its proven impact on firm performance, future research could also be targeted into sectioning R&D under different categories and examining if any particular area of R&D (AI investment), could be a potential opportunity for investment strategies. Overall, this study sets the stage for future research into the role of R&D in portfolio performance and the trade-offs between equal and value-weighted methodologies.

APPENDIX

Exhibit #1: Number of Unique Firms by Year

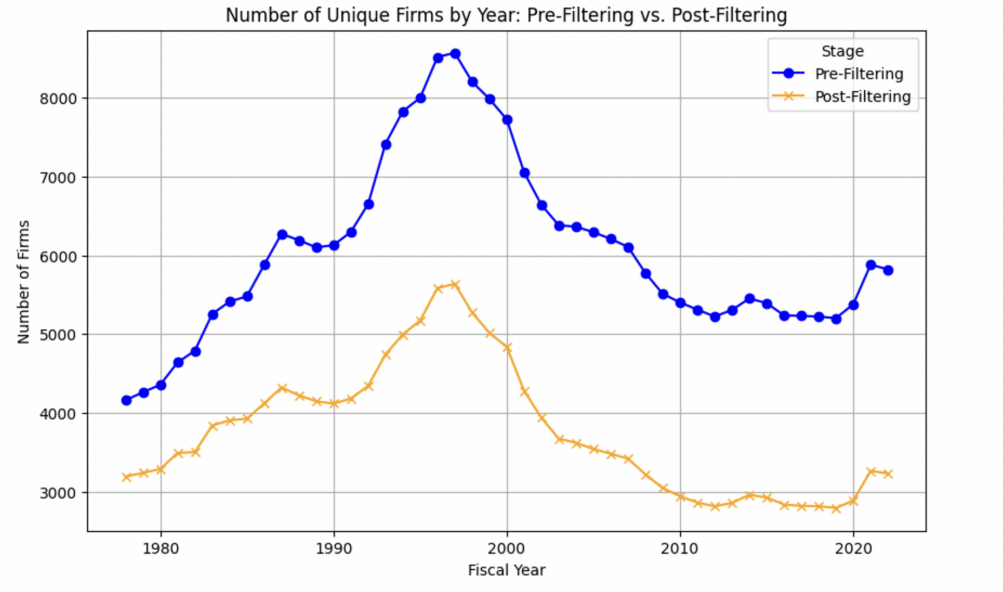


Exhibit #2: Alpha Box Plot

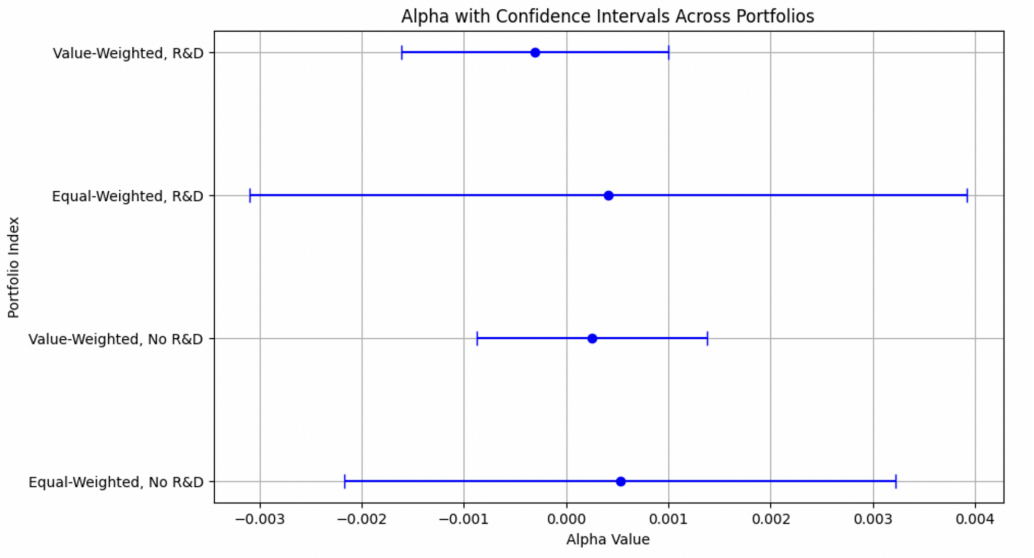


Exhibit #3: Beta Box Plot

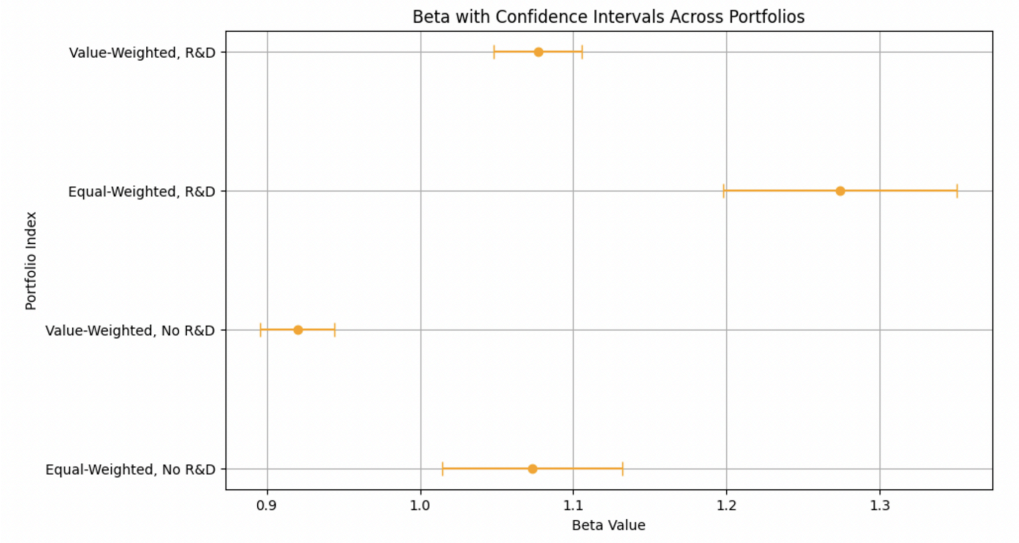


Exhibit #4: Linear Regression Plots

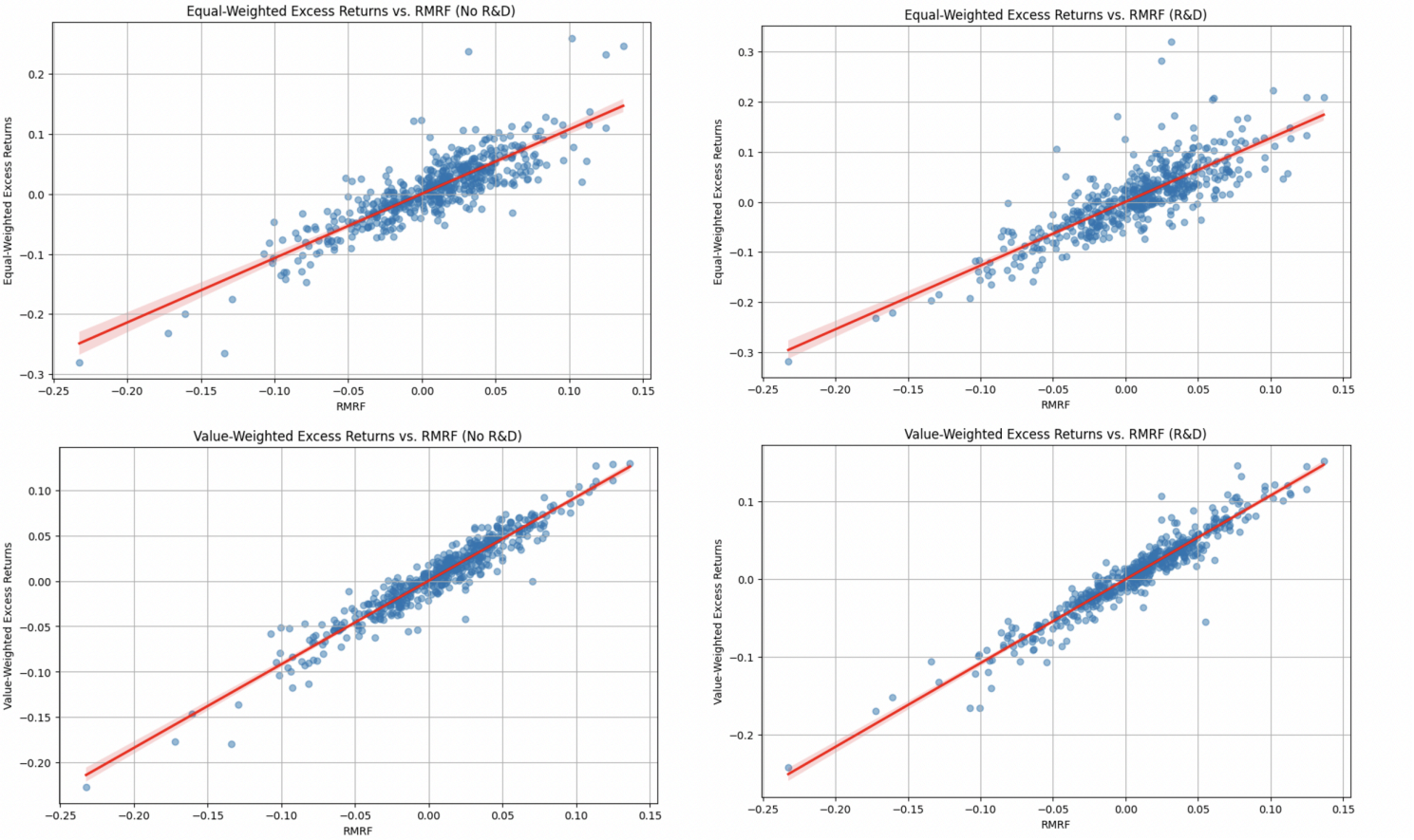


Exhibit #5: Total Market Cap by Year

REFERENCES

1. Wharton Research Data Services Library. https://wrds-www.wharton.upenn.edu/
2. Kenneth R. French Data Library. https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
3. ChatGPT. *Used during building code specifically for merging, variable shifting, and portfolio construction.*