

Investment Intensity and the Cross-Section of Future Returns

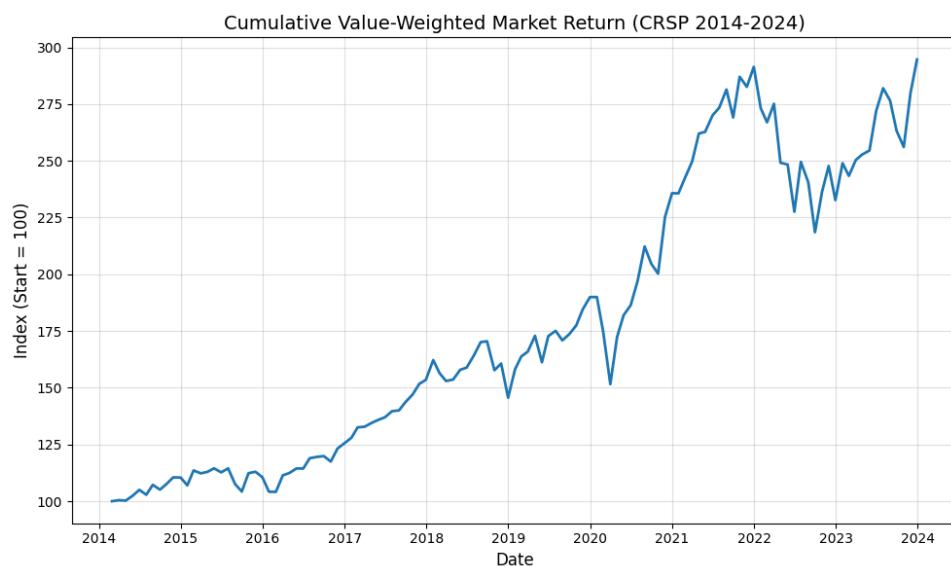
Introduction

In this report, I use data from CRSP and Compustat to examine the relationship between intangible investment and future stock returns. The analysis also explores how industry classification and firm size may influence the predictive power of both intangible and physical investment intensity.

I first describe the data construction process, then present and discuss the results corresponding to Questions 2 and 3 from the exam paper. The Appendix provides additional information, including the time spent on each section, general feedback on the exam paper, and complete regression result tables.

Data Construction

I used the CRSP–Compustat link table to merge firm-level data from both sources. The cumulative value-weighted market return from 2014 to 2024 is shown below. As the figure indicates, market returns fluctuated over time but exhibited an overall upward trend. A sharp decline occurred between 2022 and 2023, which may have been driven by the lingering economic effects of the COVID-19 pandemic.



The following table reports summary statistics for firm-month observations in the merged dataset. The sample spans the period from 2014 to 2024, and all variables are winsorized at the 1st and 99th percentiles to mitigate the influence of outliers.

Variable	N	Mean	Std. Dev.	Min	Median	Max	Units
Returns	455826	0.003	0.151	-0.426	0.001	0.579	Decimal
Price	458634	38.544	57.379	0.35	18.74	363.733	USD
Total Assets	439755	7606.379	24254.733	3.914	847.433	183010	Millions of USD
Capital Expenditures	437985	183.576	601.162	0	11.081	4317	Millions of USD
R&D Expanse	249988	135.104	489.912	0	17	4067	Millions of USD
Book Equity	442008	2021.218	6009.086	-596.5	259.658	43528	Millions of USD

Question 2: Intangibles

In this section, I discuss whether firms that devote a larger share of their resources to intangible investments tend to gain a better future return.

Regression Design

To measure how much firms devote their resources to intangible activities, I define intangible investment intensity as intangible expenditures divided by total assets.

Intangible expenditures are measured as the sum of research and development (R&D) expenses and selling, general, and administrative (SG&A) expenses:

$$\text{IntanIntensity}_{i,t} = \frac{(R\&D_{i,t} + SG\&A_{i,t})}{\text{TotalAsset}_{i,t}} \quad (1)$$

I then use an OLS regression to examine whether intangible investment intensity predicts firms' future stock returns:

$$\text{Return}_{i,t+1} = \alpha + \beta \text{IntanIntensity}_{i,t} + \gamma X_{i,t} + \delta_t + \delta_{industry} + \epsilon_{i,t+1} \quad (2)$$

Where:

- $\text{Return}_{i,t+1}$ is the future return, defined as the cum-dividend total return in the next month
- $X_{i,t}$ represents control variables, including firm size, book-to-market ratio, and momentum:
 - $\text{FirmSize} = \log(\text{MarketEquity}_{i,t})$
 - $\text{BooktoMarket Value} = \text{BookEquity}_{i,t}/\text{MarketEquity}_{i,t}$
 - $\text{Momentum} = \prod_{k=2}^{12} (1 + R_{i,t-k}) - 1$
- δ_t captures year fixed effects, controlling for market-wide shocks such as macroeconomic fluctuations or events like pandemics that affect all firms
- $\delta_{industry}$ captures industry fixed effects based on SIC codes, which help account for industry-specific shocks, also adjust for the fact that some industries naturally require higher R&D spending. Because SIC classifications can be overly detailed, I would group them into broader industry categories to make fixed effects more meaningful.

Data

a) Missing Data

For missing values in R&D and SG&A, I choose to drop those observations rather than impute zeros. It is difficult to distinguish whether a missing entry reflects the absence of such expenses or simply non-disclosure. Filling them with zeros could therefore underestimate intangible investment. Although this approach reduces the sample size, the dataset is sufficiently large that the loss of observations should not materially affect the results.

b) Industry Classification:

Intangible Group	SIC Range	Industries
High Intangible	60-67,	Finance, Insurance, Real Estate;
	70-89	Services
	20-39	Manufacturing
Medium Intangible	50-59	Wholesale Trade, Retail Trade
	15-17	Construction
Low Intangible	01-09,	Agriculture, Forestry, Fishing;
	10-14,	Mining;
	40-49, 91-99	Transportation & Public Utilities; Public Administration

Result and Discussion

The following table shows the result of OLS regression:

<i>Specification</i>	OLS Regression					
	Next Month Return					
<i>Dep. Var</i>	(1)	(2)	(3)	(4)	(5)	(6)
Intangible Intensity	-0.003 *** (-2.9556)	-0.004 *** (-3.7614)	-0.0031 *** (-2.9758)	-0.0041 *** (-3.7938)	0.0032 *** (2.5071)	0.0021 * (1.7851)
<i>Fix Effects</i>						
Time	Yes			Yes		Yes
Industry		Yes		Yes		Yes
<i>Control Variables</i>						
Firm size,						
Book to Market Value,					Yes	Yes
Momentum						
R-Square	0.0001	0.0123	0.0001	0.0123	0.0066	0.0188
N	194881	194881	194881	194881	194881	194881

Note: Robust t-statistics are reported in parentheses.

The result is surprising and interpreting the regression results as causal is challenging. Although the coefficients on intangible investment intensity are statistically significant, their signs vary across specifications. When controlling only for time and industry effects, intangible intensity is negatively associated with future returns. However, after adding firm-level controls such as size, book-to-market ratio, and momentum, the relationship turns positive. This inconsistency suggests that the observed correlation may not represent a true causal effect.

Several factors may drive this issue. First, there is a timing mismatch between variables: stock returns are measured monthly, while accounting variables such as R&D and SG&A are reported annually. As a result, it is difficult to precisely align each month's return with the underlying level of intangible investment. Even if monthly data on intangibles were available, the impact of such investments often materializes over several

years rather than immediately, making it hard to capture their short-term effects on returns. Second, omitted variable bias is likely. Stock prices and returns are influenced by many unobserved factors, such as investor expectations, market sentiment, and macroeconomic shocks, that are not fully captured by the model. Firms that invest heavily in intangibles may also have higher growth opportunities or risk exposures, which could affect their returns. Finally, reverse causality may exist. Firms performing well or with high recent returns might increase intangible spending due to managerial optimism or easier access to financing.

Together these challenges make it difficult to interpret the estimated relationship between intangible investment intensity and future returns as causal.

Empirical Strategy Proposal

a) Type of Variation

A key source of identification would come from within-firm variation over time that is driven by external factors. For example, changes in government R&D subsidies, tax incentives for innovation, or industry-level technological shocks. These shocks affect firms' incentives to invest in intangibles but are plausibly independent of short-term return fluctuations.

Controlling for firm and time fixed effects could further isolate the impact of intangible spending from unobserved, time-invariant firm characteristics and market-wide shocks.

b) Potential Instrumental Variable

One possible instrument for intangible investment is regional or industry-level R&D tax credit intensity or innovation policy exposure. Such policy variation is relevant because it directly influences intangible spending decisions. Also is exogenous because local tax policy changes are unlikely to directly affect a firm's future stock returns other than through their impact on intangible investment.

What's more, peer firms' intangible intensity within the same industry (excluding the firm itself) could serve as an instrument, reflecting common technological opportunities rather than firm-specific performance.

c) Additional Data

If additional data were available, several variables could strengthen causal interpretation:

1. Monthly or quarterly measures of intangible investment.
2. Innovation success measures, to capture realized outcomes of intangible spending.
3. Investor sentiment indicators, to control for how expectations shape current valuations.

d) Implementation

If policy or industry-level data on R&D incentives can be merged in, I would employ a two-stage least squares (2SLS) framework:

First stage: Regress intangible intensity on the chosen instrument(s) and controls.

Second stage: Regress future returns on the predicted component of intangible intensity.

(919 words)

Question 3: Heterogeneity by industry & size

In this section, I discuss whether firms that devote a larger share of their resources to intangible investments tend to gain a better future return

Classification

a) Industry classification:

NAICS codes are very granular at the sector level, so I collapse them into broad production groupings:

Industry Type	NAICS Code	Sector Title
Primary Industry, Construction and Manufacturing	11-33	Agriculture, Mining, Utilities, Construction, Manufacturing
Trade and Transport	42-49	Wholesale, Retail, Transportation
Knowledge Service	51-56	Information, Finance, Real Estate, Professional Services
Public Service	61-100	Education, Health, Leisure, Public Administration

These buckets reflect meaningful differences in the role of tangible versus intangible capital while preserving sample size for stable comparisons.

b) Firm Size Classification

The firm size is considered in three variants: market value (ME), book equity (BE) and total asset (AT). A composite size score is given by average of the percentile ranks of log (ME), log (BE), and log (AT) to smooth measurement noise. The percentiles are recomputing each month, that group membership can change over time. To avoid forcing nearly identical firms onto different sides of a single cutoff, I create a narrow median band each period and sort firms within month: Small: percentile < 48th; Median band: 48th – 52nd percentile; Large: percentile > 52nd.

Regression

To measure how much firms devote their resources to tangible activities, I define physical investment intensity as capital expenditures scaled by total assets, mirroring the definition of intangible intensity:

$$Phys_{i,t} = \frac{CapitalExpenditures_{i,t}}{TotalAsset_{i,t}} \quad (3)$$

I first estimate a pooled regression to test the average relationship between future returns and both types of investment across all firms:

$$R_{i,t+1} = \alpha + \beta_1 Intan_{i,t} + \beta_2 Phys_{i,t} + \gamma X_{i,t} + \delta_t + \delta_{industry} + \epsilon_{i,t} \quad (4)$$

Next, I use interaction models to test for systematic differences across groups. I begin by separately examining group effects for industry and firm size, then combine both dimensions if the sample size allows:

$$R_{i,t+1} = \alpha + \beta_{1,j}(Intan_{i,t} \times D_{Ind,j}) + \beta_{2,k}(Intan_{i,t} \times D_{Ind,k}) + \gamma X_{i,t} + \delta_t + \epsilon_{i,t+1} \quad (5)$$

$$R_{i,t+1} = \alpha + \beta_{1,j}(Intan_{i,t} \times D_{Size,j}) + \beta_{2,k}(Intan_{i,t} \times D_{Size,k}) + \gamma X_{i,t} + \delta_t + \epsilon_{i,t+1} \quad (6)$$

Finally, to capture potential heterogeneity along both dimensions jointly, I estimate a cross-interaction model:

$$\begin{aligned} R_{i,t+1} = & \alpha + \sum_{j,k} \beta_{1,jk}(Intan_{i,t} \times D_{Industry=j} \times D_{Industry=k}) \\ & + \sum_{j,k} \beta_{2,jk}(Phys_{i,t} \times D_{Industry=j} \times D_{Industry=k}) + \gamma X_{i,t} + \delta_t + \epsilon_{i,t+1} \end{aligned} \quad (7)$$

For comparison, I also apply the baseline specification (4) within each subgroup to examine how the relationships differ across industries and firm-size categories:

$$R_{i,t+1}^{(g)} = \alpha^{(g)} + \beta_1^{(g)} Intan_{i,t} + \beta_2^{(g)} Phys_{i,t} + \gamma X_{i,t} + \delta_t + \epsilon_{i,t} \quad (8)$$

Result

The table below reports the results of the pooled OLS regression and the interaction regressions. For the interaction and subgroup regressions, coefficients are reported as weighted averages across groups. Complete results are provided in the Appendix. Although the signs of the coefficients vary across specifications, all estimates are close to zero, suggesting that controlling for industry type and firm size does not materially alter the predictive power of intangible and physical investment intensity.

Dep.Var	Next Month Return				
	Equation (4)	Equation (5)	Equation (6)	Equation (7)	Equation (8)
Intangible Intensity	-0.003 *** (-4.675)	-0.004 * (-1.65)	0.006 (1.29)	-0.001 (-0.18)	0.001
	-0.034 *** (-2.773)	-0.009 (-0.38)	0.008 *** (3.97)	0.000 (0.03)	-0.010
Control Variables	Yes	Yes	Yes	Yes	Yes
Industry Interaction		Yes			
Size Interaction			Yes		
Full Cross Interaction				Yes	
Subgroups					Yes

Note: Robust t-statistics are reported in parentheses.

Let us now dive into the subgroup regression results. As shown below, while the coefficients on intangible and physical investment intensity vary across groups, their magnitudes remain close to zero, consistent with the pooled regression results. This suggests that neither industry classification nor firm size leads to substantial differences in the predictive power of investment intensity. The finding is somewhat unexpected, as one might intuitively expect stronger heterogeneity across industries and firm sizes.

Interestingly, compared with the results from Question 2, where only intangible intensity was considered, the explanatory power does not appear to change meaningfully. This suggests that adding physical intensity and interaction terms does not substantially improve the model's ability to explain future returns.

<i>Industry</i>	<i>Size</i>	β <i>Intangible Intensity</i>	β <i>Physical Intensity</i>	<i>R-Square</i>	<i>N</i>
Primary Industry, Construction and Manufacturing	Small	-0.0019*	-0.0341	0.0007	69266
	Median	0.0048	-0.0274	0.0052	4235
Trade and Transport	Large	0.0082***	-0.0146	0.004	51568
	Small	-0.0171*	-0.0167	0.0077	5891
	Median	0.0020	0.0505	0.0006	738
Knowledge Service	Large	-0.0088	0.0215	0.0006	10861
	Small	-0.0009	0.0199	0.0032	17628
	Median	0.0048	0.2421	0.0077	1764
Public Service	Large	0.0134***	0.0317	0.001	18095
	Small	-0.0151**	-0.0775*	0.0036	5393
	Median	0.0670*	0.1345	0.0069	704
	Large	-0.0129	0.0124	0.0016	5592

Discussion

How could the potential correlation between industry type and firm size affect the interpretation?

Because firm size and industry are potentially correlated, it becomes difficult to separate whether the effect of intangible investment reflects firm scale or industry composition. For example, large firms are sometimes concentrated in high-intangible industries such as technology and finance. This overlap complicates interpretation of the size and industry interaction coefficients, as they may capture overlapping sources of variation.

While my cross-interaction specification partially addresses this by including both industry and size effects simultaneously, the estimates could still be driven by unbalanced sample weights, as some industry-size combinations dominate the data. To check robustness, I would reweight the sample to equalize the contribution of each industry-size group and run stratified regressions within industries to isolate within-industry size effects.

(816 words)