LUISS

Degree Program in Global Management and Politics

Course of Financial Reporting and Performance Measurement

Drivers of Profitability in the Global Semiconductor Industry

An In-depth Analysis of the Semiconductor Sectors

Prof. Jonathan Berkovitch
SUPERVISOR

Prof. Luca Giustiniano
CO-SUPERVISOR

762871
CANDIDATE

Academic Year 2023/2024

Table of Contents

Abstract

This Master Thesis analyzes profitability drivers in the global semiconductor industry, using financial statements obtained from the Refinitiv database on 119 companies across integrated producers, design, manufacturing, materials, and equipment sectors from 2014 to 2023. It assesses Return on Assets through the impact of gross profits, fixed costs and asset utilization. The findings underscore sector-specific insights that would allow managers to maximize profitability: focusing on Research & Development Expenditure and Intangibles in design, prioritizing Gross Profit and Plant, Property, and Equipment in production, and managing Gross Profit with General Administrative and Marketing Expenses in integrated producers, materials and equipment providers.

Keywords: Semiconductor Industry, Profitability Drivers, Return on Assets, Sector-Specific Analysis, Fabless, Foundry, Integrated Device Manufacturer, Semiconductor Manufacturing Equipment and Services, Financial Statement Analysis, DuPont Model, United States, East Asia, Europe

1. Introduction

An integrated circuit, commonly referred to as a semiconductor (SC)[1] is the core element in the chip production for smartphones, computers, medical diagnostic or surgical equipment. These components exert a progressively substantial impact on the way people live their lives affected by the computing and electronics world[2]. The operations within the global SC market recorded sales of $570 billion in 2022 and are anticipated to exceed $1 trillion by 2030 (McKinsey & Company, 2024; Wai-Chung Yeung et al., 2023).

In addition, continued progress in the SC industry is enabling the development of new technologies in various other industries: virtual and augmented reality, the Internet of Things (IoT), device and process automation, robotics, artificial intelligence (AI), and autonomous driving.

One thread of financial analysis of a company's activity is an assessment of profitability. A company's management assesses the profitability indicators to identify the most efficient use of various types of resources, assets and capital of the enterprise (Shaker Sultan, 2014). Return on Assets (ROA) is a crucial financial metric that measures a company's efficiency in utilizing its assets to generate profits (Nithin, 2023). It serves as a key indicator of corporate success, allowing strategists and investors to compare a firm's performance against industry averages and competitors (Oliver, 2001). Therefore, it is critical form decision-makers, investors and stakeholders of the SC industry to understand the variables that influence ROA in order to make informed decisions.

The purpose of this paper is to explore the drivers of profitability in the global SC industry through an in-depth examination of its various sectors. By taking into consideration factors such as Research and Development (R&D) expenditure, non-current assets structure, and gross profit, this study identifies key elements that influence operating income and offers strategic insights for stakeholders to enhance their market position.

[1] The semiconductor industry includes the entire process of developing materials to manufacturing electronic chips (Van Zant, 2014). Within this field, the integrated circuit industry creates chips that integrate multiple electronic components into a single piece (Hu, 2024). Often, the term "microchip industry" is used interchangeably with the integrated circuit and SC industry, typically referring to the production of specific types of chips such as central processing units and graphic processing units (Ho Yeap & Javier Sayago Hoyos, 2020). Thus, while all these terms closely related to each other, the SC industry has the widest scope, the integrated circuit industry narrows to chip fabrication, and the microchip industry generally keeps the technology behind computing processors.

[2] Nowadays, a single smartphone has far more computing power than the computers used by National Aeronautics and Space Administration (NASA) to land a man on the moon during the Apollo 11 mission in 1969.

This Master Thesis (MT) proceeds as follows: Section 2 gives an overview of the SC industry. Section 3 conducts a comprehensive literature review and explains the theoretical framework. Section 4 outlines the research methodology. Section 5 presents a detailed analysis of the data and discusses the key findings. Section 6 employs Ordinary Least Squares (OLS) regression to analyze the various drivers influencing profitability across sectors and regions in the SC industry. Through multiple linear regression, this section clarifies the relationship between selected variables and their impact on profitability metrics. Finally, Section 7 concludes and provides limitations of the analysis as well as suggestions for future research.

2. Overview of the Semiconductor Industry

SC manufacturers typically organize operations around two main processes: design and manufacturing. Companies that do not have their own SC facilities and focus solely on design are named Fabless companies, while those that focus only on manufacturing are called Foundry companies. In turn, those chipmakers that do both design and manufacturing are labelled Integrated Device Manufacturers (IDM). Foundries and IDMs rely on the materials and equipment provided by Semiconductor Manufacturing Equipment (SME) producers (Wai-chung Yeung et al., 2023).

Separation by specializations has advantages. Fabless companies get high flexibility and speed of innovation. On the other hand, contract manufacturers form a pool of orders for five or more years in advance. Having guaranteed contracts from all regions of the world, Foundry-manufacturers can afford to invest in the development of production facilities and continuous improvement of the technical process (Wai-chung Yeung et al., 2023).

Furthermore, a great influence on industry is where SCs are actually produced. In 2019, six hubs of the global economy — the United States of America (USA), Europe, mainland China, South Korea, Japan, and Taiwan — accounted for approximately 92% of the global SC value chains (European Commission, 2022; Wai-chung Yeung et al., 2023).

The SC shortage which occurred in 2020 has shown that the absence of a single critical chip can prevent the sale of a device worth tens of thousands of dollars (Stewart et al., 2021). According to an analysis by Applied Energy Systems, the automotive industry suffered losses of more than $210 billion in 2019. (Applied Energy Systems, 2023). The evolution of hybrid and electric vehicles (EVs) toward greater automation has dramatically increased the number of chips found in these modern cars. This pattern shows no sign of slowing down, with the cost of microchips currently reaching $1,400 per vehicle. (Applied Energy Systems, 2023).

The widespread adoption and expansion of AI, IoT, and machine learning technologies are set to continue, finding and scaling new opportunities within the SC market. Deloitte's 2024 Global SC Industry Outlook emphasizes this trend, forecasting that the market for chips will accelerate the training and inference of generative AI models and will remain robust and is predicted to exceed $50 billion in sales for 2024 (Deloitte, 2024) This represents about 8.5% of the value of all chips expected to be sold during the year. Moreover, the forecast suggests that the demand for AI chips could potentially reach $400 billion in sales by 2027. Such chips are integral to a wide range of applications, including those found in data centers, embedded or cloud-based devices, as well as peripheral and mobile devices, underlining the SC industry's critical role in powering the current digital transformation across various sectors (Deloitte, 2024).

The SC Industry Association reported a decrease in global SC sales in 2023 (-8.3%) overall. It noticed an upturn in the latter part of the year, with a positive outlook for 2024, based on a substantial increase in sales in the last quarter of 2023 (+8.4% from Q3) (Semiconductor Industry Association, 2024).

The growing demand for Gen-AI chips, with ever-increasing chip needs in a multitude of sectors, from cars to home appliances and factories in addition to phones, computers, and data centers, and significant subsidy support from governments will be able to counter-cyclical market behaviour, the current economic downturn, and the trade war between the USA and China, as well as other geo-political tensions. The following table summarizes the SWOT analysis for the SC industry, highlighting the key drivers/factors of the industry *(see Table 1)*. An in-depth SWOT analysis can be found in Appendix (Appx.) B.

Table 1: SWOT analysis for the SC industry *(see Appx. Table B1-B4)*.

Strengths	Innovation and Technological Leadership	Continuous innovation over the past decade has led to more powerful and cost-effective devices. Key players like NVIDIA, TSMC, Intel and AMD are at the forefront of tech advances.
	Diverse Applications	SCs are crucial in numerous high-growth sectors, such as AI, IoT, autonomous driving, and healthcare, expanding their market reach and potential.
	Strong Market Position	SC corporations often secure long-term contracts and investments for future growth.
	Government Support	Significant financial support from governments, such as subsidies and favourable policies under initiatives like the CHIPS Act.

(to be continued)

Table 1 *(continued)*

Weaknesses	High Capital Intensity	The semiconductor industry requires massive capital investment in technology and manufacturing facilities.
	Cyclical Nature of the Market	The industry is prone to boom-bust cycles, as seen with the recent oversupply issues, making it difficult to manage inventory and investment levels effectively.
	Dependence on Global Supply Chains	Despite efforts to localize production, the industry is still heavily dependent on a global supply chain and is easily disrupted by geopolitical tensions, trade disputes, or global economic downturns.
Opportunities	Rising Demand for AI and IoT Technologies	Increasing integration of AI and IoT across various sectors, there is a growing demand for specialized chips, which could lead to significant revenue streams.
	Expansion in Emerging Markets	Increasing digitalization and the need for advanced technologies in emerging markets present a significant opportunity for growth.
	Healthcare and IoT Sectors	These sectors are rapidly adopting more sophisticated semiconductor technologies, driving further demand.
	New Material Innovations	Research into new semiconductor materials and processes (like GaN and SiC) can lead to breakthroughs in performance and efficiency, opening up new applications.
Threats	Economic Uncertainty and Market Volatility	Economic downturns and increased market volatility can drastically affect demand and lead to significant financial challenges.
	Intense Competition	The competitive landscape, particularly with the rise of fabless companies and foundries, puts pressure on pricing, margins, and market share.
	Subsidies and Global Challenges	Government subsidies are crucial for reducing investment payback times in new SC factories, while geopolitical tensions and tech nationalization are prompting companies to focus on supply chain resilience.

Source: (Adzan et al., 2017; Counterpoint, 2024; European Commision, 2023; Gartner, 2023; Internation Roadmap for Devices and SystemsTM, 2020; KPMG & GSA, 2024; Liu et al., 2020; McKinsey & Company, 2024; Microsemi PPG, n.d.; MMR, 2023; Mordor Intelligence, 2023; Neff, 2023; SC-IQ: Semiconductor Intelligence, 2023; Shilov, 2024a, 2024b; Stewart et al., 2021; Tulyagankhodjaev et al., 2023; Wai-chung Yeung et al., 2023; Zhai, 2023)

3. Literature Review and Theoretical Framework

Financial analysis, including profitability assessment, is essential in business valuation, helping to assess a company's readiness to take financial risks and its ability to repay debts (Hristozov, 2021).

Return on Assets (ROA) is a ratio that evaluates how effective a company is in utilizing its assets to generate profits (Singh, Chaudhary, et al., 2023). Therefore, when trying to make a forecast of the financial well-being of an enterprise, many analysts focus on studying the dynamics of this indicator and its drivers. Profitability metrics are relative, as they compare various types of profit, such as gross, operating, pre-tax, and net profit, to the respective flows that generated them or to the assets utilized in earning these profits. (Rutkowska-Ziarko, 2015a). One advantage of ratios is to be able to compare companies of different sizes and also those whose financial statements are originally expressed in different currencies (Gibson, 1982).

Recent studies evaluate profitability across diverse regions and look into the financial performance of companies. Industry-specific profitability has also been associated with ratios like the fixed asset ratio, asset turnover, and sales to current assets ratio (Burja, 2011a). Close inspection of the SC industry highlights the importance of considering in detail the nominator and denominator when dealing with asset-based financial ratios. Operating Profit, proxied by Earnings Before Interest and Taxes (EBIT), will be used to ensure the focus remains on evaluating the Gross Sales Margin and efficiency of the industry's core operating activities, producing a more accurate image of the company's operational health and industry performance. As for the denominator, we will be using end-of-the-period total assets [3] *(see Appx. Section C1-C2).*

The analysis will be enhanced by employing the DuPont model to break down the drivers of profitability. The DuPont approach helps to analyze the factors that affect a company's return. Depending on the depth of the analysis, the model version used also becomes more detailed (Ram & Chouhan, 2020). The original DuPont model, introduced in 1912 for an internal efficiency report, combined Operating Margin with Asset Turnover (Jory et al., 2021). It can be summarized in equation **[1]**:

Furthermore, the formula can be expanded to include a breakdown of the Operating Margin into

$$ROA = \frac{EBIT}{Revenue} \times \frac{Revenue}{Total\ Assets} \quad \textbf{[1]}$$

two: Operating Efficiency and Gross Sales Margin. This allows for a more detailed look at the core activity efficiency and fixed versus variable cost management within the company (Ram & Chouhan, 2020). The expanded formula provides deeper insights into where profitability is being driven or hindered within the organization **[2]**:

[3] ROA can be calculated using either the average of total assets between two periods, or end-of-the-period total assets. The use average total assets smooth out fluctuations caused by asset purchases, sales, or seasonal variations (Jewell & Mankind, 2012). The average of total assets between two periods should be applied when there are fluctuations in the total assets due to purchases, sales, or industry seasonality. The end-of-the-period total assets is less accurate for business environments where total assets change significantly over the year due to periodical variations. Such a sector is the construction building industry where the cycles, in fact, are related to the growth of GDP and the cycles are repeated over and over again. In the SC industry this is not the case. The cycles incorporate a lot of new technology. Evolution of cycles follows stages, but these are never repeated. Each new stage is different from the previous one, as it will incorporate very different technologies, so there cannot possible be repeated cycles.

$$ROA = \frac{EBIT}{Gross\ Profit} \times \frac{Gross\ Profit}{Revenue} \times \frac{Revenue}{Total\ Assets} \quad \textbf{[2]}$$

Some research concentrates on exploring SC sectors' asset productivity, specifically based on the sector, namely Foundries and Fabless; (Shin et al., 2017). While broader studies are addressing the semiconductor industry as a whole, they date back to the end of the last century, and thus would not be able to fully reflect the current SC industry dynamics (Irwin & Klenow, 1994; Kozmetsky & Yue, 1998). To the best of our knowledge, no study has solely analyzed the identification of the profitability drivers across the global SC industry sectors in this century.

Many studies have focused on the impact of Research and Development (R&D) and Intangible Assets (Intg). This underlines the important role of new technologies in the value creation in the SC industry (Goodall et al., 2002; Heck & Pinner, 2007; Helms, 2003; Shin et al., 2017; Weber, 2002). Several econometric papers demonstrate the correlation between Intg and profitability across different industries (Burja, 2011a; Matiş et al., 2010). More recently negative correlation between Intg and ROA has been found by other research (Mendez-Morales et al., 2024). The impact of Intg on profitability differs across markets, and industries, and depends on the type of Intg (Tiron Tudor et al., 2014).

Plants, Property & Equipment (PPE) as a driver for performance has contradicting outcomes. A Polish study showed a slight positive correlation between non-current assets and profitability. However, for PPE, the correlation was slightly negative (Zimny, 2022). In contrast, research on certain Nigerian manufacturing firms presents a positive link between PPE and ROA (Udoayang et al., 2020). These two studies highlight the relationship between PPE and financial outcomes across different sectors and geographical locations. Another study shows that PPE and Equipment Asset Productivity are crucial for the foundry sector in the SC industry (Li, 2010). Despite this, there is still not much research in this area.

Hence, based on the industry overview and literature review, we have identified the following profitability drivers relevant to the SC sector as possibly being of interest: Gross Profit, EBIT, R&D expenses, PPE, and Intg. In order to include these industry-specific metrics, we developed a unique DuPont model to provide a more thorough picture of ROA for the SC industry. The new formula is illustrated in equation **[3]** and explained in more detail in Appx. D *(see Appx. Equation D1-D14)* breaks down the three-fold ROA **[2]** into three subsections corresponding to Operating

Efficiency, Gross Sales Margin and Asset Turnover[4]. It takes into account the complex relationship between increasing operational efficiency and growing asset utilization. It isolates the effect of R&D investments and asset management on profitability in this specific industry.

$$ROA = \left[1 - \frac{R\&D}{Gross\ Profit} - \frac{Other\ SG\&A\ exp.}{Gross\ Profit}\right] \times \left[\frac{Gross\ Profit}{Revenue}\right] \times \left[\frac{Revenue}{Curr.A} \times \left(1 - \frac{PPE}{Tot.A} - \frac{INTG}{Tot.A} - \frac{Other\ Non-Current\ Assets\ \ (TOTAL)}{Tot.A}\right)\right] \quad \textbf{[3]}$$

From the literature *(see Appx. E, Table E1)*, to the best of our knowledge this proposed approach to breaking down ROA has not been utilized anywhere else before. This model can provide an explanation of what drives profitability in the SC industry, particularly how and to what extent R&D, PPE, and intangible assets interact with other financial elements. These variables were identified as being very important in this specific industry. By adapting the DuPont model, a clearer perspective is provided including more detailed ways to assess the fundamentals of a company. The purpose of this study is to further our understanding of the global SC industry and its drivers.

4. Methodology

Research questions

This Master Thesis investigates the profitability drivers in the SC industry over the period 2014-2023. To do so, the following research question (RQ) will be answered:

RQ1: *How does the ROA evolve in the SC industry as a whole and in its sectors?*

RQ2: *Is there an association between region, sector, fiscal year and firm size in the SC industry?*

RQ3: *How do various drivers influence profitability across different sectors and regions within the SC industry?*

RQ4: *How and to what extent do these drivers impact the SC industry?*

To address RQ1, we will conduct a univariate analysis comparing the overall averages of Return on Assets (ROA) and winsorized ROA (ROA_win). We will further break down these metrics by sector and region to gain deeper insights into the data. For RQ2 a non-parametric analysis will be utilized to explore the associations between categorical variables *(see Appx. Section J1)*. To respond to RQ3, we will perform bivariate analysis and conduct ANOVA tests *(see Appx. Section K1-K2)* to assess the impact of the categorical variables on independent continuous variables, while also considering average values by sector and region. Finally, for RQ4, multiple linear

[4] Asset turnover in an average of turnover of current and non-current assets, this new formula isolates these two parts.

regression analyses will be carried out for each sector to explore how various drivers affect profitability across different sectors and regions within the SC industry.

Variables

A proforma analysis of items in the company fundamentals showed which of these have the greatest weight.[5] Relevant variables are listed in the following table, both general firm characteristics and financial ratios *(see Table 2)*. A detailed version of the table with the categories and the equations for the continuous variables, can be found in the Appx. H *(see Appx. Table H1-H3)*[6].

Table 2: *Variables*

Variable	*Acronym*
Firm characteristics (Categorical variables)	
Region	REGION
SC industry sector	SECTOR
Firm size	FMSIZE
Fiscal year	YEAR
Financial ratios (Continuous variables)	
Gross Sales Margin	GROSS
Current Asset Turnover	CAT
R&D expenses ratios	R&Dg
SG&A minus R&D ratio	SG&A*
PPE ratio	PPEa
Intangible Assets ratio	INTGa

From the industry and literature review, we conclude that investment in innovation, proxied by R&Dg, PPE and Intg, plays a critical role in the SC industry. It is worth noting that IDMs, Fabless and SME&Ss focus heavily on intellectual property, while Foundry companies will need to invest heavily in PPE in order to improve their efficiency and keep pace with the latest technologies. Tech companies frequently reinvest a substantial portion of their gross profits back into such expenses in order to drive future growth and innovation. Gross Sales Margin can directly impact the amount of resources available for such reinvestment. Additionally, due to the specialized market niches that SME&Ss operate within, their fixed costs are significantly shaped by their other Selling, General, and Administrative expenses (SG&A) such as marketing, in addition to R&D investments. Another important factor that needs to be considered in the SC industry is the current assets. There is a difference between asset-heavy and asset-light segments within the industry. Asset-heavy companies, (e.g. Foundries and SME&S), would benefit significantly more from cash reserves and inventory in terms of improved ROA. On the contrary, asset-light companies, such as Fabless, show a less pronounced impact (Lin & Huang, 2011).

[5] The proforma analysis consists of accounts benchmarked to the Revenue in the income statement and to the Total Assets in the balance sheet.
[6] Other accounts either hold no significant relevance based on the benchmarking or are not relevant to the research.

Sampling

The analysis includes companies from all the sectors in the SC industry and covers the years 2014 to 2023. The selection of publicly traded corporations guarantees full access to their financial data. Companies were chosen based on industry research and corporate rankings accessible as of February 2024 (Companies Marketcap, 2024). Analyzing such a large sample over an extended period offers a more complete picture of the industry's financial dynamics.

The data source is the Refinitiv database and supplemented with financial statements from the USA's Securities and Exchange Commission and the companies' websites to check for discrepancies (U.S. Securities and Exchange Commission, 2024).

The period chosen, 2014 to 2023 (the most recent year for which there is available data at the moment of writing the Master Thesis), is characterized by significant technological advances[7]. These innovations increased the demand for microchips in the automotive industry, IoT and AI. This boom, starting in 2014, fueled manufacturing in the SC industry, significantly affecting its economic performance. (Consumer Electronics Association (CEA), 2014; Deloitte, 2024; Stewart et al., 2021). Moreover, major global events, such as the trade war between the USA and China, and the COVID-19 pandemic altered the economic landscape during the period of analysis. It can be observed how businesses addressed these economic shocks and responded to the subsequent stages of recovery (Accenture, 2021; KPMG India, 2024).

The significance of the distribution across specific regions of the world was behind the decision to categorize the world's companies into three distinct regions for this research: the USA, Europe, and East Asia. This allows for a more structured analysis of the SC industry's global dynamics, as well as recognizing the nuanced contribution and strategic positions of these key geographical areas in the sector's worldwide ecosystem. Furthermore, in the industry analysis, five distinct sectors in the SC industry were identified, namely: IDMs, Fabless, Foundries, SMEs, Material & Service providers[8]. The sample was further segmented by company size, categorized as either "Big company" or "Mid-sized company", based on their Log winsorized Total Asset values. Detailed criteria for size classification are provided in Appx. F *(see Appx. Section F1-F3).*

---

[7] Intel created the 14nm microarchitecture in 2014, and in the following years this developed into the 7nm and 5nm microarchitectures (Intel, 2014).

[8] However, due to the considerable similarities in the business models of the latter two categories, this Master Thesis consolidates them into a single sector, Semiconductor Materials, Equipment, and Service (SME&S) providers.

The initial dataset consists of 119 sample corporations (see *Appx. Table G1*). Sixteen companies were excluded due to insufficient data for three or more years, as follows: two in Europe, four in East Asia, and ten in the USA *(see Appx. Script M1-M7)*[9]. Furthermore, six companies were excluded due to their geographical location, namely one from Australia[10], one from Canada[11] and four from Israel[12] (see *Appx. Table G1*). Three other corporations that have negative equity over several years were excluded *(see Appx. Table G1)* and four companies were excluded based on the individual fiscal year-end as some had not yet completed the fiscal year 2023 at the time of data collection *(see Appx. Table G1)*. Further exclusion was due to omission of data and/or inconsistencies between the financial statements in the 10-K reports, and the Refinitiv database[13] *(see Appx. Table G2)*. After a careful comparison of estimated and reported values, the final sample was composed of 91 companies, which included a total of 893 financial statements *(see Appx. Table M3-G6)*. The contingency table *(see Table 3)* describes the distribution of different SC sectors within the industry across the regions.

Table 3: Distribution of SC companies by sector and by region

	IDM	Fabless	Foundry	SME&S	**Total**
USA	*13*	*17*	*2*	*16*	*48*
East Asia	*5*	*9*	*7*	*8*	*29*
Europe	*3*	*3*	*1*	*7*	*14*
Total	*21*	*29*	*10*	*31*	*91*

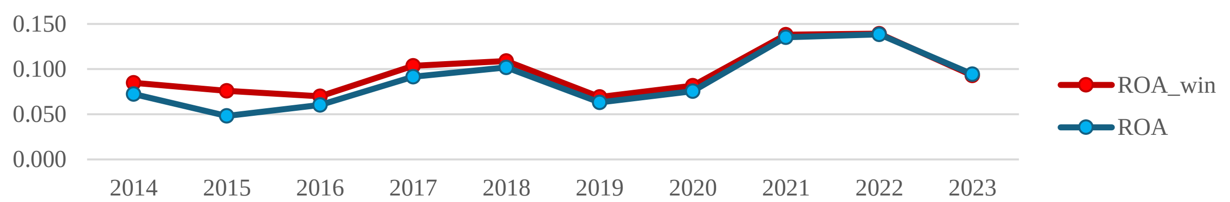
[9] Executed via Python 3 script *(see Appx. Script M1 - M7)*. Companies with only one or two years of unreported data were not excluded.
[10] BrainChip
[11] POET Technologies
[12] Nova Measuring Instruments, Tower Semiconductor, Camtek, Valens Semiconductor
[13] To guarantee the accuracy of the financial data obtained from the Refinitiv database, re-calculation was done using established financial accounting equations. It focused on checking current *versus* non-current assets and liabilities, as well as total equity. Several discrepancies were identified, which led to manual cross-analysis of the annual reports for the relevant years. Numerous errors were corrected due to divergencies with the 10-K reports. Nonetheless, 25 discrepancies remained unresolved, and thus those observation points were excluded.

5. Data Analysis

Return on Assets (RQ1)

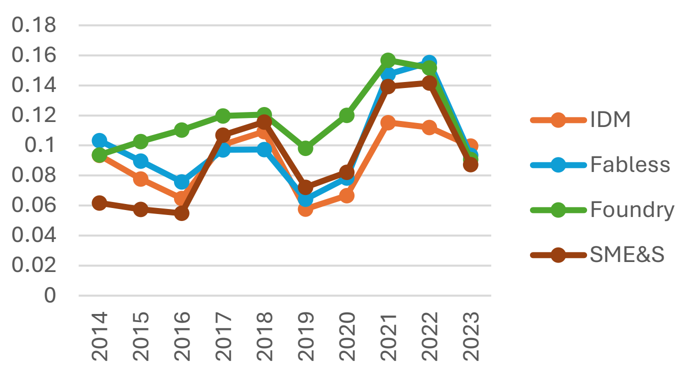
The variables show a series of upswings and downturns. In the short term, ROA swings sharply from year to year, with a steady rise between 2016 to 2018, only to reach a low in 2019, followed by a sharp recovery in the years 2021 and 2022 and a sharp fall in 2023, which aligns with findings (Semiconductor Industry Association, 2024). In the medium term, we observe four-year cycles with a slight growth trend. Over the long term, despite evolution by waves every four years in medium terms, the ROA in 2023 remains nearly the same as it was in 2014. *(see Figure 3).*

Figure 3: Return on Assets (total) of the Semiconductor industry



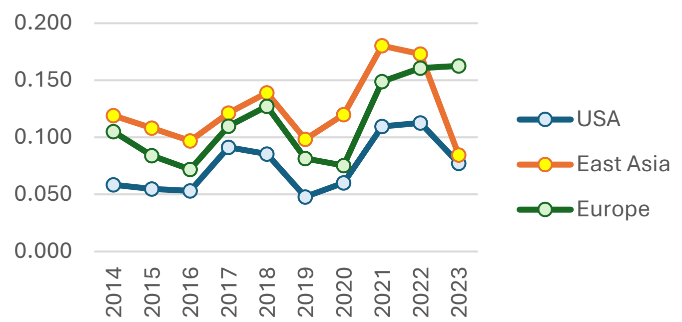
An extended overview of each sector shows that from 2018, all the sectors have the same pattern. Only Foundry differs and stays above the other sectors till recently, as 2023 seems to show a change. They all show a downturn in 2023. The IDM sector is relatively steady but does not reach the heights of others. On the other hand, Fabless firms are more erratic but manage to achieve higher returns, peaking notably in 2018 and again in 2022. Foundries start slow but pick up pace in the following years, showing a marked improvement. The SME&S sector shows a gradual, consistent rise *(see Figure 4).*

Figure 4: winsorized Return on Assets (by sector)



When we look at ROA by region, the USA shows quite a bit of variability, hitting a low in 2015 but steadily climbing back up through 2023. East Asia stands out for its consistently high ROA, particularly after 2016, indicating strong profitability until its rapid decline in 2023. Nevertheless, there is a dramatic fall in 2023, and it ends lower than in 2014. Europe had a good year in 2023, in contrast to other regions and throughout the period of analysis shows a big increase and continues to rise steadily *(see Figure 5).*

Figure 5: winsorized Return on Assets (by region)



Independence/association between the categorical variables Firm characteristic (RQ2)

After performing six non-parametric, chi-square tests[14] to check for an association between the categorical variables *(see Appx. Section J1)*, the following results were identified:

YEAR is not correlated to any of the three variables. It shows no significant association with the FMSIZE, the SECTOR of the company, nor the REGION *(p = 1.0E+00,* for all three tests*)*. Size matters a little when it comes to association with region as FMSIZE and REGION are associated but not strongly *(p = 3.6E-02)*. Meanwhile, FMSIZE and SECTOR are significantly associated *(p = 1.3E-33)*.

Region and sector are correlated. The results show that the SECTOR is quite strongly associated with the REGION *(p = 1.1E-16)*. This suggests that the sector a company operates in is strongly influenced by the region it is located in. It could imply that certain industries are concentrated in specific geographical areas, possibly due to local resources, market demands, or policy incentives. ANOVA[15] indicates that REGION and SECTOR significantly affect all ROA drivers, while the impact of FMSIZE on these drivers varies and is often less significant *(see Appx. Section K1-K2)*.

---

[14] Execution via Python script *(see Appx. Script M9)*

[15] Executed via Python 3 script *(see Appx. Script M10-M11)*

Profitability drivers (RQ3)[16]

In Appx. K the following tables can be found on which the analysis was conducted: Table K1, detailing mean financial metrics across different sectors in the semiconductor industry; Table K2, presenting these metrics by region; and Table K3, combining both sector and region to provide detailed financial metrics *(see Appx. Table K1-K3).*

H_1* *1: *There is a significant difference between* ***GROSS*** *across (a) the Sectors and (b) the Regions.*

The analysis shows that USA and East Asian Foundries exhibit high Gross Sales Margins at *0.427* and *0.342* respectively. This suggests efficient operations and market strategies, while European corporations have a Gross Sales Margin of *0.157*. The USA Fabless companies lead with a higher margin of *0.551*, in front of Europe and East Asia *(0.446* and *0.453* respectively*)*, possibly indicating higher competitive pressures in Europe. The IDM, Fabless and SME&S sector shows a Gross Sales Margin of *0.422, 0.511, and 0.399* respectively, with all three regions being fairly uniform and close to the mean. Region-wise, the USA reports the highest Gross Sales Margins *(0.464)*, followed by East Asia with *0.409*, driven by robust market performance and possibly lower operational costs. Europe shows the lowest margins *(0.382)*, which may reflect higher costs or intense competition. This being the case, ***H_1 2 (a)*** and ***H_1 2 (b)*** are considered valid.

H_1* *2: *There is a significant difference in* ***CAT*** *based on (a) the Sector and (b) the Region.*

The IDM sector leads in current asset turnover with a rate of *1.520*, indicative of a highly efficient use of assets to generate revenue. The Fabless, Foundry and SME&S sectors show more moderate *CAT*, suggesting different operational rates and/or market conditions. In terms of regional performance, European fabless lead with *1.690* over their USA and East Asia counterparts *(1.290* and *1.170* respectively*)*. Hence, we cannot reject ***H_1 2 (a)*** and ***H_1 2 (b)***.

H_1* *3: *There is a significant difference in* ***R&Dg*** *based on (a) the Sector and (b) the Region.*

Based on the Analysis the highest R&D spending on average across the regions is presented by the Fabless sector *(0.500)* followed by IDM companies *(0.344)* and lowest by the SME&S *(0.278)* and the Foundry sector *(0.270)*. In the IDM sector, the R&D expenses are led by Europe *(0.379)* and the USA *(0.368)*, showing a strong commitment to innovation, while East Asia presents lower rates of *0.260*. Foundries show the least focus on innovation. Europe tops R&D investment *(0.573)*,

[16] Given the big sample size employed, the mean provides a comprehensive average that is appropriate and efficient for large datasets, leading to a good understanding of central tendencies across the variables. The values presented here span the entire period of analysis, in order to provide consistent assessment of trends and differences.

emphasizing technological advancement, followed by East Asia *(0.253)* while the USA invests the least *(0.169)*. The Fabless sector, which has the highest innovation focus, sees the USA leading in R&D spending *(0.551)*. East Asia and Europe also invested significantly *(0.349* and *0.453)*, indicating robust R&D activities. For SME&S, the USA again leads in R&D investment *(0.329)*, with Europe following *(0.278)*, and East Asia coming last *(0. 172)*, reflecting different regional goals. Hence, we cannot reject ***H_1 3 (a)*** and ***H_1 3 (b)***.

H_1 4: *There are significant differences in* **SG&A*** *based on the (a) Sector and (b) Region.*

The SME&S also maintains the highest SG&A* position *(0.404)*, reflecting their strategic emphasis on marketing and administration to support their operations and growth, while IDM, Fabless and Foundries have smaller expense ratios of *0.325*, *0.330* and *0.252*. Notable are the European Foundry companies show a strikingly high ratio of SG&A* at *0.574* compared to USA *(0.169)* and European *(0.253)* Foundries. When it comes to regional differences. The USA is leading with a striking *0.422*, followed by European companies *(0.286)* and East Asian firms *(0.243)*. Hence, we cannot reject ***H_1 4 (a)*** and ***H_1 4 (b)***.

H_1 5: *There is a significant difference in* ***PPEa*** *based on (a) the Sector, and (b) the Region.*

When we look at the PPE assets across sectors, European *(0.227)* and East Asian *(0.257)* Foundries show significant investment followed by the USA *(0.157)* totalling to average ratio of *0.214*. This is closely followed by IDM's average at *0.270*, with East Asia's dominating ratio of *0.372* over the USA *(0.241)* and Europe's *0.227*. Meanwhile, the Fabless and SME&S sectors show more modest investments with *0.102* and *0.193*, respectively. Fabless in Europe have a ratio of *0.148* while the USA and East Asia have almost equal *0.096* and *0.099*. In the SME&S sector, PPE investment is led by European and East Asian SME&S with *0.233* and *0.245* respectively, while the USA has *0.152*. This trend underscores a heavy reliance on physical infrastructure in the Foundry sector compared to others. Regionally, East Asia leads the way in PPE asset investment in the IDM sector with a *0.372* ratio, highlighting its focus on maintaining robust physical operations. The USA and Europe follow with ratios of *0.241* and *0.227*, respectively. In contrast, the Fabless sector maintains the lowest PPE assets across all regions, with the USA and East Asia displaying almost negligible figures, reinforcing the sector's lean operational model. Hence, we cannot reject ***H_1 5 (a)*** and ***H_1 5 (b)***.

H_1 6: *There is a significant difference in* ***INTGa*** *assets based on (a) the Sector and (b) the Region*.

In the IDM sector, the Intg assets are led by Europe *(0.323)* and the USA *(0.284)*, showing a strong commitment to intellectual property, while East Asia presents lower rates of *0.100*. Foundries show the least focus on intellectual property. In Intg the USA is the undisputed leader *(0.137)*, while Europe and East Asia trail behind *(0.010* and *0.013* respectively*)*. In the Fabless sector, which has the highest innovation focus, in Intg, we can see that East Asia *(0.065)* trails significantly behind the West *(0.275* for the USA and *0.178* for Europe*)*. For SME&S, the USA again leads in Intangibles *(0.164)*, with Europe following closely *(0.112)*, and East Asia coming last *(0.044)*, reflecting different strategic goals. Hence, we cannot reject ***H_1 6 (a)*** and ***H_1 6 (b)***.

H_1 7: *There is a significant difference in* ***winsorized ROA*** *based on (a) the Sector, (b) the Region.*

The Fabless East Asian and European sectors are also doing well, with an ROA of *0.139* and *0.153* respectively, and a surprisingly low ROA_win from the USA at *0.068*. Similar is the situation when it comes to SME&S. East Asia and Europe lead with *0.137* and *0.112* respectively while the USA trails behind with *0.064*. The IDM sector has lower returns, at *0.089* with all three regions being very close to the average. Looking at the regions, East Asia tops the list with an ROA_win of *0.127*, likely due to efficient operations and strong market performance. The USA, on the other hand, has the lowest at *0.075*, which could mean higher costs or tougher competition are cutting into profits. Europe is in the middle with an ROA_win of *0.112*, showing a steady performance across its businesses. Hence, we cannot reject ***H_1 7 (a)*** and ***H_1 7 (b)***.

6. Regression

Research models for multiple linear regression analysis

The regression analysis of economic indicators involves many metrics. Via in-debt analysis, the model can assess how the distribution and use of different indicators such as a company's capital, resources, and sales or services rendered impact the efficiency of an enterprise's management and profitability (Dumbravă, 2010; Veblen, 1908). This MT explores how and to what extent different elements such as gross profit, PPE assets, research spending, and others impact a company's ROA by applying multiple linear regression using Ordinary Least Squares *(see Appx. Section L2)*. The research is based on panel data for ten years and covers 91 companies belonging to four sectors of the SC industry. For each sector, one pooled OLS regression is estimated. The model is unique, but it is applied to four SC sectorial samples: IDM, Fabless, Foundry and SME&S. Additionally, dummy variables are used for firm size and region. One dummy variable distinguishes between big firms and mid-sized firms (FMSIZE, a categorical variable which assumes the value 0 when a company is Big, and otherwise, 1). The other two dummy variables compare the regions of East Asia and Europe against the USA (REGION). To account for the significant role that the sector plays in the SC industry (SECTOR) adjusts the analysis to factor in sector-specific variations. By doing so, it can be seen more clearly how each factor works in its own industry context. This sector-based approach guarantees that the unique traits of each business model are considered and do not skew the overall results, increasing the robustness of the findings. The index ***i*** represents the different observation points and ***j*** is used to distinguish between the different model specifications, in this case SECTOR (j = IDM, Fabless, Foundry, SME&S).

$$ROA\_win_{ij} = \beta_{0j} + \beta_{1j}GROSS_{ij} + \beta_{2j}CAT_{ij} + \beta_{3j}R\&Dg_{ij} + \beta_{4j}SG\&A^{*}_{ij} + \beta_{5j}PPEa_{ij} + \beta_{6j}INTGa_{ij} + \beta_{7j}(\text{East Asia} - \text{USA})_{ij} + \beta_{8j}(\text{Europe} - \text{USA})_{ij} + \beta_{9j}FMSIZE_{ij} + \epsilon_{ij}$$

The following hypotheses are tested via the regression models:

H₁ 9: *GROSS has a statistically significant impact on ROA_win across the sectors.*

H₁ 10: CAT *has a statistically significant impact on ROA_win across the sectors.*

H₁ 11: R&Dg *has a statistically significant impact on ROA_win across the sectors.*

H₁ 12: *SG&A* has a statistically significant impact on ROA_win across the sectors.*

H₁ 13: *PPEa has a statistically significant impact on ROA_win across the sectors.*

H₁ 14: INTGa *has a statistically significant impact on ROA_win across the sectors.*

The regression models *(see Table 4)* have strong explanation capacity, as demonstrated by their relatively high Adj. R^2. The statistical significance of the F-tests, with all p-values below 0.001, confirms the models' general validity. The dummy variables *(East Asia - USA)* and *(Europe - USA)*, have mixed impacts across sectors. The dummy *(East Asia – USA)* has a significant positive effect in the Fabless sector, hence when it comes to ROA_win, firms in this region perform differently to their USA counterparts *(see Appx. Table L12)*. The assumption tests suggest some issues with normality and mild autocorrelation concerns in some groups, with generally manageable multicollinearity in the dataset (*see Appx. Section L4*). All variables were standardized, for easier comparison of impact magnitudes across predictors within each model (*see Appx. Section L5*).

Table 4: Regression results	**Model-IDM**		**Model-Fabless**		**Model-Foundry**		**Model-SME&S**	
N (Number of data points)	*198*		*282*		*94*		*289*	
Adj. R^2	0.714		0.638		0.899		0.768	
F-test (p-value)	< .001		< .001		< .001		< .001	
Predictor	**Stand. Estimate**	**p**	**Stand. Estimate**	**p**	**Stand. Estimate**	**p**	**Stand. Estimate**	**p**
Independent variables								
GROSS	0.564	< .001	0.368	< .001	0.818	< .001	0.327	< .001
CAT	0.150	0.003	0.320	< .001	0.346	< .001	0.254	< .001
R&Dg	-0.184	0.003	-0.413	< .001	-0.036	0.577	-0.322	< .001
SG&A*	-0.340	< .001	-0.070	0.270	-0.027	0.680	-0.457	< .001
PPEa	0.059	0.427	0.031	0.512	-0.379	< .001	-0.184	< .001
INTGa	-0.343	< .001	-0.372	< .001	-0.189	< .001	-0.199	< .001
Dummy variables								
East Asia – USA	0.012	0.906	0.466	< .001	0.309	0.030	-0.014	0.873
Europe – USA	0.144	0.263	0.475	< .001	0.271	0.194	0.101	0.227
Mid-sized firm – Big firm	-0.506	< .001	-0.355	< .001	-0.064	0.559	-0.177	0.014

Interpretation (RQ4)

All models are positively affected[17] by ***GROSS***, with a moderate impact on Fabless and SME&S. Fabless primarily spends on design and engineering, which are less variable than manufacturing costs. SME&S face intense competition and price sensitivity. The impact is more substantial on the IDM and Foundry sectors. High Gross Sales Margins in IDMs are essential to cover high fixed costs from their extensive infrastructure. Foundries require significant capital expenditure to maintain and upgrade facilities, ensuring that Gross Sales Margins are adequate to sustain operations in a changing demand environment and technological advancements. Therefore, ***H_1 9*** cannot be rejected for all four models.

CAT has a medium positive impact on all models. IDM and SME&S have lower ***CAT***, while Fabless and Foundries have slightly higher ***CAT***. The lower ***CAT*** in IDMs can also reflect longer production cycles and inventory holding periods, given the complexity of managing an entire production from design to manufacture. Similar to IDMs, SME&S may also have a significant portion of their assets in non-current forms, particularly if they are engaged in manufacturing or

[17] ROA increases when the gross sales margin increases.

provide capital-intensive services. Fabless companies typically have fewer fixed assets and more liquid assets than the other sectors. Foundries though are involved in manufacturing like IDMs. They often operate with a business model optimized for high efficiency. Those findings align with the conclusions drawn in the literature review (Chang & Wu, 2022; Tsai & Chou, 2015). Thus, ***H_1* *10*** cannot be rejected for models Fabless, Foundry and SME&S but for Model-IDM.

R&Dg has the strongest negative impact on Fabless, followed by SME&S, while IDM and Foundries have lower, with Foundries' impact being close to zero with a p-value indicating that the variable is not significant for them. They focus on production technology and efficiency rather than product development. The result for Fabless is expected to rely on innovation for a competitive edge. SME&S can include a variety of businesses from design to niche manufacturing services, and their R&D needs can vary significantly. But in their case the outcomes do not quickly generate revenue, leading to a negative impact on financial performance. IDMs can amortize R&D costs over a larger base of operations and products The impact is lower compared to Fabless and SME&S because their broader operational scope can better absorb and justify R&D expenses. This appears to be consistent with the findings in the literature review (Goodall et al., 2002; Kozmetsky & Yue, 1998). As a result, ***H_1* *11*** is rejected for models IDM and Foundry, but cannot be rejected for models Fabless and SME&S.

SG&A* mainly includes marketing and administrative expenditure, and also has a negative impact. The strongest impact is presented in the IDM and SME&S regressions while the variable is not significant for the Fabless and especially for Foundries. IDMs need robust administrative and sales infrastructures to manage their vast integrated activities. SME&S often operate on a smaller scale with potentially less efficient economies of scale compared to larger companies. Their SG&A expenses include efforts to market and administrate across possibly diverse but niche markets. The non-significant impact on Fabless and Foundries reflects more streamlined operations with less need for expansive sales and administrative structures. Their main differentiation lies in innovation (for Fabless) and production (for Foundries) rather than extensive marketing or administrative functions. ***H_1* *12*** cannot be rejected for models IDM and SME&S but for models Fabless and Foundry.

PPEa has a negative impact on SME&S and a strong negative impact on Foundries. For IDMs and Fabless, the variable is strongly insignificant. SME&S, depending on their specific roles within the SC industry, might require substantial investments in equipment and facilities, particularly if they are engaged in manufacturing or testing. Foundries are heavily reliant on sophisticated and expensive manufacturing equipment to produce SC products. The strong negative impact here suggests that while necessary, the high costs associated with maintaining and upgrading such assets weigh significantly on their ROA. This could be due to rapid obsolete tech requiring frequent updates. The insignificance of IDMs suggests that their scale and integration efficiently offset the costs associated with PPE through enhanced productivity and control over production processes. Fabless investment in PPE is minimal, as they do not own manufacturing facilities, hence their capital is not tied up in heavy equipment, which matches the evidence presented by the literature review (Li, 2010). That is why ***H_1 13*** is rejected for models IDM and Fabless, and cannot be rejected for models Foundry and SME&S.

INTGa has a medium negative effect on the IDMs and Fabless and a smaller effect on the Foundries and SME&S. Intangible assets for IDMs can include intellectual property such as patents, software, and proprietary technologies crucial for both product development and manufacturing processes. Fabless also heavily depend on intangible assets like design patents and software. These are vital for maintaining a competitive edge. Foundries may have investments in intangible assets such as process technology and manufacturing know-how. However, their intangible assets are more about enhancing efficiency rather than product development. SME&S firms include a variety of roles, including specialized manufacturing, services, or design. This appears to be consistent with the findings in the literature review (Park et al., 2021; Shin et al., 2017). Their intangible assets could range widely depending on their specific niche, from patents and technologies to brand value. To conclude, ***H_1 14*** cannot be rejected for all four models.

7. Conclusion

The purpose of this MT was to research the key drivers of ROA in the SC industry. By adopting a customized DuPont model, uniquely developed for this research, we analyzed financial data of the 91 largest, publicly traded SC companies over a span of ten years (2014-2023). It focuses on factors that highlight the differences in ROA within the industry across regions of the world (USA, Europe and East Asia) and between the various sectors, namely the IDM sector (integrated designs and manufactures of their own chips), Fabless companies (chip designers who outsource manufacturing), Foundries (the outsourced production factories used by Fabless), and SME&S companies (providers of the materials, tools, services and testing). Sector-specific factors: Selling, Administrative, R&D, Intangible and PPE expenses, have varying degrees of influence on ROA.

The evolution of ROA over time shows how important the regional dynamic is. The ROA of SC companies in East Asia outperforms those from Europe and the USA regions in the period under analysis. In addition, cyclical behaviour of profitability in the industry is observed, with market volatility and the recent downturn in 2023. Regression analyses reveal that Gross Sales Margin (GROSS) is a significant driver of ROA in the SC industry, though its positive impact varies across sectors, with the most significant effect among the Foundries. Furthermore, specific variables identified in the SWOT analysis as key variables in the SC industry significantly affect their ROA. R&D intensity has a significant and negative impact across sectors, being stronger in Fabless and SME&S than in IDM and Foundry sectors, this effect suggests that investments in innovation, although necessary, have a big toll on ROA in the short term[18]. Moreover, there is a strong negative impact of PPE on Foundries, a sector where companies are highly dependent on sophisticated, and high-cost fabrication assets. Finally, intangible assets (INTGa), like patents, have a moderate negative effect on IDMs and Fabless, albeit higher than in Foundries and SME&S. This research addresses academic and industry analysts, as well as decision-makers in the SC industry, especially those who wish to better understand the financial dynamics in SC production around the globe. It may also be of interest to investors and managers who need to be aware of the effects on profitability levels of decisions taken about strategic ventures and operational efficiencies in this capital-intensive industry.

The main contribution of this research is its comprehensive global analysis of profitability across SC sectors, expanding the understanding of the financial metrics that drive expenditure and earnings. It adds to the existing literature by considering both geographical and operational factors.

Limitations and suggestions for future research

Data bias is evident, as certain sectors are underrepresented in some regions (e.g. only one Foundry company is represented in Europe). It is also worth noting that companies from Canada (1), Israel (4) and Australia (1) were excluded from the study, due to geographic criteria. By analyzing them separately in a country-specific study and including them in future research, might provide a more comprehensive understanding of the SC. Future research could focus on exploring the impact of geopolitical factors or the macroeconomic drivers of profitability, such as GDP

[18] ROA is a short-term financial ratio that, when analyzed over multiple periods, can reveal trends. Including macroeconomic variables like taxation levels and interest rates – which vary by region – can provide further insights. Additionally, incorporating variables such as GDP growth can enhance our understanding of economic cycles in the medium term (Qolbi et al., 2020).

growth rates, which might help to understand the cyclical evolution of ROA. The long-term effects of government subsidies and regulations (e.g., the CHIPS Act) on sectoral profitability and competitiveness, could be researched, particularly in regions like the USA and Europe.

There is also a correlation between some independent variables, such as R&Dg and SG&A*, and PPEa and INTGa, all fixed assets. This may lead to multicollinearity, which can affect the model by distorting the estimated effects of the independent variables[19].

This Master Thesis did not include non-financial drivers of ROA, yet ESG factors are increasingly impacting investment decisions, potentially affecting ROA. Therefore, there is motivation to explore the association of environmental sustainability initiatives and ROA and their financial implications for the SC industry.

[19] Regularization methods, such as Ridge or Lasso regression, can help by adding a penalty to the model for increased complexity (i.e., using many or highly correlated predictors). Ridge regression is particularly good at dealing with multicollinearity by shrinking the coefficients of correlated predictors (Abdulhafedh, 2022).

Appendices

Appendix L: Multiple linear regression models

Section L1	Relations of economic variables
Section L2	Method of Ordinary Least Squares (OLS)
Section L3	Analysis of the quality of the sample equation of multiple linear regression
Section L4	Assumptions check
Section L5	Standardization
Table L1	IDM correlation matrix
Table L2	Fabless correlation matrix
Table L3	Foundries correlation matrix
Table L4	SME&S correlation matrix
Table L5	Model-IDM regression results
Table L6	Model-Fabless regression results
Table L7	Model-Foundry regression results
Table L8	Model-SME&S regression results
Table L9	Assumption check – Normality test
Table L10	Assumption check – Autocorrelation test
Table L11	Assumption check – Collinearity Statistics
Table L12	Dummy variables detailed interpretation

Appendix M: Python and R Scripts

Script M1	Extracting Company names and Headquarters
Script M2	Extract paths and register
Script M3	Extract Statement data and Period end date
Script M4	Extract variables (Income statement)
Script M5	Extract variables (Balance sheet)
Script M6	Detection of companies with more than two years of missing statements
Script M7	Exclude specified companies
Script M8	Winsorization of variables
Script M9	$\chi 2$ test of independence/association
Script M10	ANOVA - Analysis of Variance
Script M11	Kruskal-Wallis, Levene's test, and Shapiro-Wilk test - Tests for Assumptions and Distribution

Appendix A: Abbreviations and acronyms

Table A1: List of abbreviations and acronyms used in the Master Thesis

Abbreviation	Explanation
$	United States Dollars
AI	Artificial Intelligence
AMD	Advanced Micro Devices
Appx.	Appendix
CAT	Current Asset Turnover
CHIPS act	Creating Helpful Incentives to Produce Semiconductors act
EBIT	Earnings Before Interest and Taxes
EVs	Electric Vehicles
FMSIZE	Firm size
GROSS	Gross Sales Margin
H_1	alternative hypothesis
IDM	Integrated Device Manufacturer
INTGa	Intangible Assets ratio
Intg	Intangible Assets
IoT	Internet of Things
MT	Master Thesis
NASA	National Aeronautics and Space Administration
OLS	Ordinary Least Squares
PPE	Plants, Property & Equipment
PPEa	PPE ratio
R&D	Research and Development
R&Dg	R&D expenses ratio
REGION	Region
ROA	Return on Assets
ROA_win	winsorized ROA
RQ	Research Question
SC	Semiconductor
SECTOR	SC industry sector
SG&A	Selling, General, and Administrative expenses
SG&A*	SG&A minus R&D ratio
SME	Semiconductor Manufacturing Equipment & Services
SME&S	Semiconductor Materials, Equipment, and Service
TSMC	Taiwan Semiconductor Manufacturing Company Limited
USA	United States of America
YEAR	Fiscal year

Appendix B: In-depth SWOT analysis

Table B1: Detailed Exploration of Strengths from the SWOT Analysis

Innovation and Technological Leadership
Counterpoint Research highlighted NVIDIA's and AMD's growth due to AI deployments as a standout in 2023, in the middle of the oversupply situation. Those firms forecast AI, memory sector recovery, and automotive sector growth as primary drivers for the SC market in 2024 (Counterpoint, 2024). TSMC is the undisputed leader in terms of scale, complexity, and reach. The company produces millions of wafers annually for major customers in virtually every sector of the semiconductor industry. The Taiwanese company has spent more than three decades perfecting the manufacturing process which has made it a technology leader in the field (Wai-chung Yeung et al., 2023). The classic integrated manufacturer is the American company Intel. Because of its mixed model of semiconductor manufacturing, it has great momentum both in designing and developing innovations, as well as in updating production facilities and developing the technological process. Furthermore, integrated manufacturers such as Samsung, in addition to contract manufacturing services, also design chips for their own needs and under their brand (Wai-chung Yeung et al., 2023).
Diverse Applications
McKinsey & Company expects Global SC chip production to grow at a compound annual growth rate of 68% estimated to expand to around $1 trillion by the end of 2030 *(see Figure B1)*. SCs will remain critical across all industries. Key sectors, such as EVs, are expected to see strong growth in demand, which could multiply with new applications for autonomous driving and electromobility. Wireless communication, computing, and data storage will also make a significant contribution to the development of the industry (McKinsey & Company, 2024). In healthcare, growing SC spending was expected to boost the development of remote monitoring and workflow processes. A significant number of healthcare devices depend on semiconductors; in fact, at least 50% of medical devices incorporate them. This reliance on semiconductors is a trend that continues to grow stronger (Zhai, 2023).
Strong Market Position of Key Players
Deloitte highlighted the significant impact of chip shortages on the automotive industry, leading to a substantial increase in the cost of microchips per vehicle—from $300 in 2010 to over $500 in 2022 (Stewart et al., 2021). This rise, amounting to more than $60 billion in microchip costs for the year, shows cars' increasing digitization and highlights the automotive sector's high dependence on SC producers (Stewart et al., 2021). Another indicator of the SC industry's recovery is Micron's recent announcement that the company has successfully pre-sold its entire 2024 production of the new DRAM[20] memory chip and has already allocated its 2025 production. This is a notable turnaround, considering that memory chips were significantly oversupplied in 2023 (Shilov, 2024b).
Government Support
A standout development is the collaboration between the U.S. Department of Commerce and Intel, which will lead to a significant deal under the CHIPS and Science Act, securing for Intel $8.5 billion in funding, and access to $11 billion in low-interest loans, plus a notable tax credit for investments. (Shilov, 2024a). Furthermore, in early 2022, the USA President approved the U.S. Innovation and Competition Act, including the Chip Act, under which the USA government allocated $52 billion for the development of domestic SC production, research and development. Similar budget support measures are also being taken in Europe, Japan, South Korea, and other countries, but the largest share of new capacity is expected to be created in the coming years in the USA, China, Japan and Europe (European Commission, 2023; Wai-chung Yeung et al., 2023).

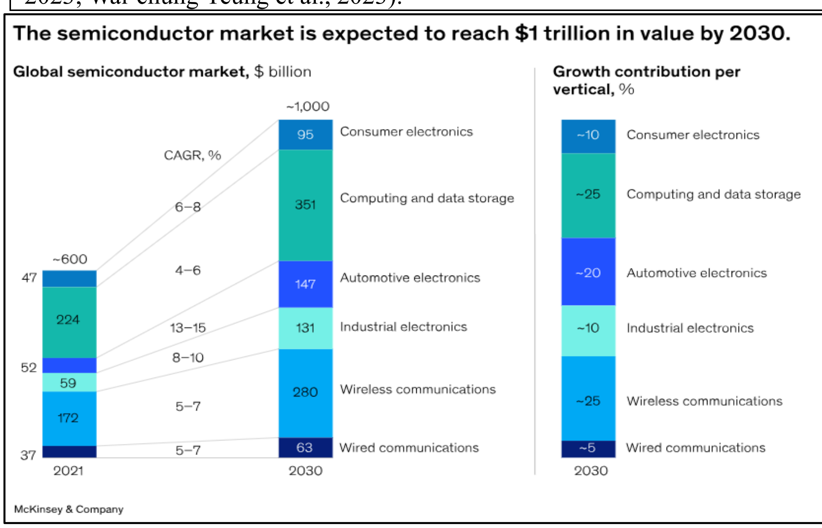


Figure B1: Semiconductor market expectations by 2030

Data source:
SC-IQ: Semiconductor Intelligence, 2023

[20] NAND and DRAM are two types of memory technologies used in electronic devices. NAND memory, often found in flash storage devices, is a type of non-volatile storage. This means it does not require power to retain the data stored on it. DRAM (Dynamic Random Access Memory), on the other hand, is a type of volatile memory, which means it loses its data when power is turned off.

Table B2: Detailed Exploration of Weaknesses from the SWOT Analysis

High Capital Intensity
At the end of the second quarter of 2023, the year was seen as a year of significant decline, with market falls expected to be between 15% and 20%, signalling a possible reduction in capital investment as companies become cautious in reaction to market turmoil (SC-IQ: Semiconductor Intelligence, 2023). This expected decline underlines the industry's ongoing challenge of balancing aggressive expansion during upswing periods with fluctuations in demand and leads to a cautious approach to investment in the face of imminent market fluctuations.
Cyclical Nature of the Market
Gartner provides a comprehensive analysis indicating that the SC market entered a correction cycle due to the demand downturn for personal computers, smartphones, and other consumer electronics, leading to an oversupply of NAND and DRAM memory chips. By the end of 2022, the SC inventory began oversupplying, particularly in the memory market, due to weakened demand as the economic slowdown (growing interest rates, and high inflation). The situation was reflected in revenue recorded for 2023 and forecasts for 2024, with Gartner recording a significant worldwide decline in SC revenue in 2023 (-10.9%) (Gartner, 2023).
Dependence on Global Supply Chains
In the face of ongoing global challenges such as trade disputes between the USA and China, conflicts in Ukraine and the Middle East, and a trend towards the nationalization of key technologies, SC companies are prioritizing the resilience of their supply chains (KPMG & GSA, 2024). This is because SC supply restrictions are negatively affecting the electronics and automotive industries around the world. Negative trends have intensified the efforts of various countries to develop and create their own SC capacities "on their own soil", as a result of which they are directing tens of billions of dollars of budget investments for this purpose.

Table B3: Detailed Exploration of Opportunities from the SWOT Analysis

Rising Demand for AI and IoT Technologies
Based on the insights gathered, it is reasonable to conclude that advancements in microchip technology are expected to evolve in close alignment with developments in AI, suggesting an interdependency between the two fields *(see Figure B2)*. AI chips are specialized semiconductors that incorporate AI technology and are used for machine learning. The need for more efficient systems to solve mathematical and computational problems has become critical as the volume of data increases. In addition, the emergence and mass adoption of autonomous robotics is expected to form a potential near-vertical growth opportunity for the AI chips market. According to Maximize Market Research, the AI chip market, which was estimated at \$21.73 billion in 2023, is projected to reach US\$202 billion by 2030. This projection shows growth at a compound annual growth rate of +37.5% over this period (MMR, 2023).
Expansion in Emerging Markets
One of the fastest-growing segments could be the automotive sector. As mentioned above, it represents a significant opportunity for SC growth, with the market expected to grow from \$71.62 billion in 2023 to \$140.52 billion by 2028, fueled by the integration of advanced technologies and the commitment to achieving zero emissions. Asia will be the fastest-developing region in this sector, driven by growing vehicle production and increasing demand for electric cars (Mordor Intelligence, 2023).
Advancements in Healthcare and IoT Sectors
SC demand in applications such as healthcare, smart cities, and consumer technology (generally named IoT) is growing and is continuously expanding. Smart city initiatives rely extensively on the IoT, necessitating increasingly advanced, compact, and affordable chips. Each new innovation demands an ever-growing number of chips to function (Adzan et al., 2017). By 2025, it is projected that over 74 billion IoT devices will be installed worldwide, a number expected to soar to 125 billion by 2030. Each of these devices will require semiconductors to operate, underscoring the critical role semiconductors play in enhancing urban efficiency and connectivity (Internation Roadmap for Devices and SystemsTM, 2020).
New Material Innovations
Upcoming innovations in SC materials and manufacturing technologies have the potential to significantly change the industry landscape. Advances such as miniaturization of chip geometries are greatly sped up by sophisticated manufacturing techniques, largely enabled by developments in AI. This miniaturization requires forming thin parts and placing them on a wafer at the nanoscale, with additional difficulty added by the metals used to reduce interconnect latency. This allows chip manufacturers to fabricate components more cost-effectively and quickly. These advances in microchip processing contribute to advances in robotics in their turn (Liu et al., 2020). SC R&D is focusing on new materials, such as gallium nitride (GaN) and silicon carbide (SiC), to push the boundaries of chip miniaturization. These materials are selected for their capacity to be crafted into smaller sizes, endure higher temperatures, support greater voltages, and switch more rapidly than conventional materials (Microsemi PPG, n.d.). In late October 2023, a groundbreaking SC material named 'Re6Se8Cl2' was discovered[21]. This innovation could possibly open new horizons for the future of SC technology (Neff, 2023).

[21] This discovery provides with unprecedented efficiency (Tulyagankhodjaev et al., 2023). The material is based on the flow within a quantum framework, allowing acoustic exciton-polarons to outpace electrons in silicon. Remarkably, this advancement could increase processing speeds to femtoseconds, six orders of magnitude beyond the currently achievable gigahertz electronics' nanosecond capabilities (1fs = 1e-6 ns), and all achievable at room temperature.

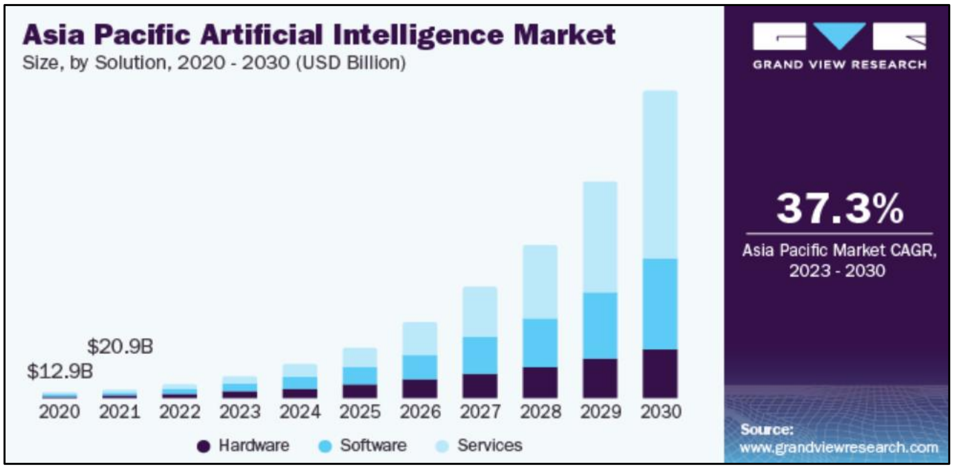


Figure B2: Projection of AI market development

Data Source:
(GVR, 2023)

Table B4: Detailed Exploration of Threats from the SWOT Analysis

Economic Uncertainty and Market Volatility
The dynamics of semiconductor capital investment and market growth reflect a pattern of rapid investment in boom years followed by significant market rises *(see Figure B3)*. Since 1984, significant market growth in excess of 20% has been followed by an increase in capital investment. For example, the expanding period in 2021 saw a 34% upswing in capital investment, led by companies such as TSMC, which reported a stunning 74% increase to US $30 billion. However, this sort of behaviour often leads to over-capacity, which triggers a downturn when demand declines (SC-IQ: Semiconductor Intelligence, 2023)
Intense Competition
Most of the key industry players are focusing on the development of AI chips and applications. NVIDIA Corporation, Advanced Micro Devices, Intel Corporation, NXP semiconductors, Analog Devices, Qualcomm Incorporated, MediaTek, Microsemi Corporation, South Korean giant Samsung Electronics, and a successful AI startup, privately held Graphcore, are all engaged in this area.
Subsidies and Global Challenges
Government subsidies will play a crucial role in the SC industry, influencing the building of new manufacturing factories *(see Figure B4)*. The size of these subsidies also correlates with reduced payback periods for these capital-intensive ventures. McKinsey & Company analysis reveals that higher subsidies are needed to substantially reduce the time it takes for fabs to pay back investment, with the effect of this financial support varying from region to region (McKinsey & Company, 2024). In the face of ongoing global challenges such as trade disputes between the USA and China, conflicts in Ukraine and the Middle East, and a trend towards the nationalization of key technologies, SC companies are prioritizing the resilience of their supply chains (KPMG & GSA, 2024).

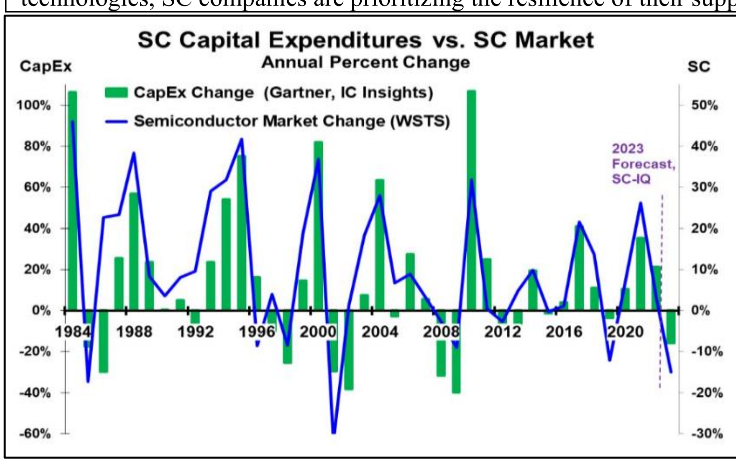


Figure B3: Semiconductors Capital Expenditures versus Semiconductors Market

Data source:
SC-IQ: Semiconductor Intelligence, 2023

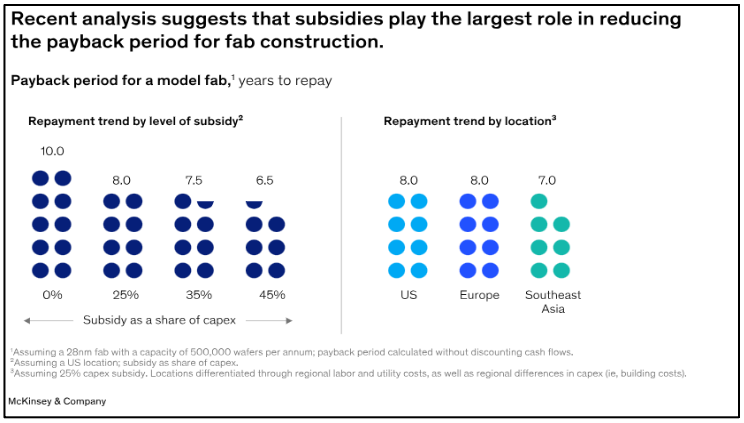


Figure B4: Role of Subsidies in the payback period for fabs construction

Data source:
McKinsey & Company, 2024

Appendix C: Choice of Nominator & Denominator for ROA

Section C1: *Nominator*

For the ROA calculation, several indicators of its profitability can be used in the numerator: Net Income (NI) or Profit after Tax (Higgins, 2009; Ross et al., 2022), the sum of NI and interest expense (Berk & DeMarzo, 2020), and others. Some authors defend that ROA should be calculated based on the pre-tax income, that is, Earnings Before Taxes (EBT), thus excluding the effect of taxation or the EBIT on Total Assets (Filatov, 2018; Ivanova, 2019).

This Master Thesis uses the modern method proposed by Singh, Gupta and Chaudhary (2023), that is, the EBIT when calculating the ROA of the enterprise (Singh, Chaudhary, et al., 2023). One reason for using EBIT comes from its applicability in the financial analysis of the semiconductor industry. The operating profit is unaffected by non-operating items, debt levels, taxes, or dividends (Jewell & Mankind, 2012). This feature makes it important when measuring operational performance, as it is free of external financial structures.

Following examination of the initial descriptives of the semiconductor sector from 2014 to 2023, non-operating income and financial investments are negligible to the performance of the companies within the sector, as non-operating income typically accounts, on average, -1% of revenue, long-term investments account for around 2% of total assets, and investments in associates, joint ventures, and unconsolidated subsidiaries account for 1% of total assets.

Hence, compared to the return on operating activities, the financial activities return is negligible. This has determined the decision to use EBIT as the numerator in the ROA calculation. This approach ensures the profitability and efficiency of the industry's core operating activities are the focus of evaluation, which will produce a more accurate image of its operational health and industry performance.

Section C2: *Denominator*

ROA can be calculated using either the average of total assets between two periods, or end-of-the-period total assets. The usage of average total assets smooths out fluctuations caused by asset purchases, sales, or seasonal variations (Jewell & Mankind, 2012). The average of total assets between two periods should be applied when there are fluctuations in the total assets due to purchases, sales, or industry seasonality. The end-of-the-period total assets are less accurate for business environments where total assets change significantly over the year due to seasonal variations. Such a sector is the construction building industry where the cycles, in fact, are related to the growth of GDP and the cycles are repeated over and over again. In the SC industry, this is not the case. The cycles incorporate a lot of new technology. The evolution of cycles follows stages, but these are never repeated. Each new stage is different from the previous one, as it will incorporate very different technologies, so there cannot possibly be repeated cycles.

Appendix D: Development of expanded ROA formula

Equation D1: Original DuPont for ROA

$ROA = \frac{EBIT}{Revenue} \times \frac{Revenue}{Total\ Assets}$	[D1]

Equation D2: Expanded three-step DuPont for ROA (Ram & Chouhan, 2020)

$ROA = \frac{EBIT}{Gross\ Profit} \times \frac{Gross\ Profit}{Revenue} \times \frac{Revenue}{Total\ Assets}$	[D2]

Equation D3 – D4: EBIT first breakdown

$EBIT = (Revenue - Variable\ Costs) - Fixed\ Costs$	[D3]
$EBIT = Gross\ Profit - Fixed\ Costs$	[D4]

Equation D5 – D7: Fixed Costs breakdown

$Fixed\ Costs = R\&D\ expense + Labor\ \&\ Related\ expense + Administrative\ expense\ (unclassified) + Other\ Operating\ expense$	[D5]
$Fixed\ Costs = R\&D\ expense + (Selling, General\ \&\ Administrative\ expense - R\&D\ expense)$	[D6]
$Fixed\ Costs = R\&D + SG\&A^{*}$	[D7]

Equation D8: EBIT second breakdown

$EBIT = Gross\ Profit - R\&D - SG\&A^{*}$	[D8]

Equation D9 – D10: Asset Turnover first breakdown

$Asset\ Turnover = \frac{Revenue}{Current\ Assets} \times \frac{Total\ Assets - Non.current\ Assets}{Total\ Assets}$	[D9]
$Asset\ Turnover = \frac{Revenue}{Current\ Assets} \times \left(1 - \frac{Non.current\ Assets}{Total\ Assets}\right)$	[D10]

Equation D11 – D12: Non-current Assets breakdown

$Non.current\ Assets = Investments\ (Long\&Term) + Investments\ in\ Associates\ ... + Receivables\ \&\ Loans + Derivative\ ... + Property\ Plant\ \&\ Equipment + Assets\ Held\ for\ Sale + Intangible\ Assets + Other\ Non.Curr.Assets$	[D11]
$Non.current\ Assets = PPE + Intangible\ Assets + Other\ Non.current\ Assets$	[D12]

Equation D13: Asset Turnover second breakdown

$Asset\ Turnover = \frac{Revenue}{Current\ Assets} \times \left(1 - \frac{PPE + Intangible\ Assets + Other\ Non.current\ Assets}{Total\ Assets}\right)$	[D13]

Equation D14: Expanded ROA that includes the breakdowns

$ROA = \left(1 - \frac{R\&D}{Gross\ Profit} - \frac{SG\&A^{*}}{Gross\ Profit}\right) \times \frac{Revenue}{Current\ Assets} \times \left(1 - \frac{PPE + Intangible\ Assets + Other\ Non.current\ Assets}{Total\ Assets}\right)$	[D14]

Appendix E: Literature Review on SC Industry and ROA

Table E1: Overview of past research about drivers of profitability in the SC industry and ROA

Objectives	Authors & Year	Period of Analysis	Industry	Region	Sample	Dependent Variable(s)	Independent Variable(s)	Main Findings
"R&D and Firm Performance in the Semiconductor Industry"	(Shin et al., 2017)	2000 - 2010	SC industry: Fabless & IDMs	Worldwide	11 IDMs & 10 Fabless.	Sector	"Gross Margin, Net Margin, ROA, Fabless (dummy), R&D ratio, Size, Capital intensity"	Fabless firms, which outsource production, boast higher margins, superior returns on assets, and greater intangible value than IDMs. Both firm types exhibit similar relationships between R&D spending and returns on assets or intangible value. However, fabless firms experience a more pronounced negative impact of R&D spending on net margins, as they allocate a larger share of sales to R&D. Despite this, the effect of R&D spending on return on assets and intangible value shows no significant difference between fabless and integrated firms.
"The Impact of Technological Capability on Financial Performance in the Semiconductor Industry"	(Park et al., 2021)	2005 - 2016	SC industry	Global - prelevance to US market	92 SC firms: 51 Fabless & 41 IDMs	Revenue, Market Cap	"Technological capability, Technological intensity/diversity, Technological asset, Technological efficiency"	Firms with high technological assets benefit more from technological intensity in terms of financial performance, whereas those with low technological assets gain more from technological diversity. Conversely, companies with high technological efficiency experience greater financial gains from technological diversity, while those with lower technological efficiency see better financial outcomes from technological intensity.
"Best Practice: An Optimization of Assets Productivity in Semiconductor Manufacturing"	(Li, 2010)	n.a	SC industry: Foundries	n.a	4 Foundries	Asset Productivity, Equipment Asset Productivity	"Average selling price, Billing loading, Total Assets per piece, PPE per piece"	The paper introduces the Equipment Asset Productivity (EAP) metric to enhance the assessment of asset management across semiconductor fabrication plants, complementing the existing Asset Productivity (AP) metric. It applies these metrics to pinpoint the main asset categories and specific fabrication plants contributing to subpar asset management at the corporate level. Additionally, it analyzes Return on Equity (ROE) components—financial leverage, return on sales (ROS), and asset turnover (AT)—to diagnose the underlying causes of poor asset productivity, providing guidance for Fab directors and top management to optimize asset management performance.
"Comparative Performance of Global Semiconductor Companies"	(Kozmetsky & Yue, 1998)	1982-1994	SC industry	USA & East Asia	56 SC firms	Labour productivity, Cost efficiency, Profit margin, R&D expense ratio, Growth trend of company's market value	"Sales Revenue, Employment, Market share and economic growth"	As Japanese companies increased their market share, this growth was linked to decreasing profit margins, reduced cost efficiency, and poor stock market performance. Conversely, USA companies have maintained competitive advantages primarily through higher profit margins and better cost efficiency. The competition in the global market for SC has escalated to a point where government interventions have become necessary. The paper also highlights the significant influence of government trade policies in this context.
"Concave Effect of Financial Flexibility on Semiconductor Industry Performance: Quantile Regression Approach"	(Chang & Wu, 2022)	n.a	SC industry	Taiwan	137 publicly traded SC firms	ROA, Return on Equity	"Financial flexibility, Cash flexibility"	The study reveals that in Taiwan's semiconductor industry, the relationship between financial flexibility (FF) and firm performance (FP) forms a concave or inverted U-shape. This suggests there is an optimal level of financial flexibility that maximizes firm performance. Specifically, this concave FF-FP relationship is evident in the lower and median quantiles of the overall semiconductor industry. It is also observed in the lower quantiles of asset-heavy business model (AHBM) firms in IC-design and manufacturing, and across most quantiles of asset-light business model (ALBM) firms. However, for AHBM semiconductor companies, financial flexibility has an insignificant impact on firm performance.
"Application of Multiple Output Data Envelopment Analysis in Interpreting Efficiency Improvement of Enterprise Resource Planning in Integrated Circuit Firms"	(Tsai & Chou, 2015)	1997-2006	integrated circuit (IC) industry	Taiwan	25 integrated circuit (IC) firms	Firm efficiency measures (technical efficiency, pure technical efficiency, and scale efficiency), Number of granted patents	"GDP, Log of Market Cap, North America SC equipment book-to-bill ratio"	Implementing Enterprise Resource Planning (ERP) in IC firms not only stimulates innovation activities, as evidenced by a notable increase in granted patents, but also significantly enhances operational efficiency. This improvement in efficiency becomes apparent after the third year of ERP implementation, demonstrated by gains in technical efficiency, pure technical efficiency, and scale efficiency. These advancements in operational efficiency are primarily due to reduced inventory and accounts receivable turnover days, coupled with an increase in accounts payable turnover days.

(to be continued)

Table E1 *(continued)*

Objectives	Authors & Year	Period of Analysis	Industry	Region	Sample	Dependent Variable(s)	Independent Variable(s)	Main Findings
“Factors Influencing The Companies‘ Profitability”	(Burja, 2011b)	1999-2009	Chemical industry	Romania	1 company	ROA, Return on Equity	“Fixed Asset Ratio, Debt to Assets Ratio, Debt to Equity Ratio, Sales to Current Assets Ratio, Sales to Equity Ratio, Gross Margin Revenue on Inventory Ratio, Expense Revenue Ratio”	“For performance indicator Return on assets were identified some influence factors that through their common action can contribute to increasing or lowering of the profitability of the analyzed company.”
“Model for the Analysis of the Company Performance”	(Dumbravă, 2010)	2009	Plastic fabrication	Romania	SC DELTA SRL	ROA	“Fixed Asset ratio, Intangible Asset ratio, Tangible Asset ratio, Financial fixed assets ratio, Circulating assets ratio, Cash ratio”	Identified are the main indicators for the analysis of the patrimonial situation: the working capital, the necessary working capital, the net treasury and the net situation.
“The Influence of Profitability Ratios and Company Size on Profitability and Investment Risk in the Capital Market”	(Rutkowska-Ziarko, 2015b)	2008-2014	Food	Poland	15 companies listed on the Warsaw Stock Exchange	ROA	“ROA mean, ROA semi-deviation, ROA standard deviation, mean rate of stock return, standard deviation of a rate of return, semi-deviation of a rate of return against a risk-free rate, company size”	The profitability of sales for food companies listed on the Warsaw Stock Exchange significantly influences both the average rates of return and the risk faced by investors. This relationship serves as evidence that, at least within this industry, the dynamics of the capital market are substantially impacted by the profits that these companies generate. This insight underscores the importance of company performance in shaping investor outcomes in the financial markets.
“Defining Return on Assets (ROA) in empirical corporate finance research: a critical review”	(Singh, Gupta, et al., 2023)	2001-2021	corporate finance literature	Worldwide	100 scholarly articles	ROA	“profit after tax - PAT, EBIT, Profitability Ratio”	Overall, 66 studies have utilised the EBIT-based ROA whereas 34 have utilised the PAT-based ROA

Appendix F: Company Size Classification Methods

Section F1: *Annual classification (Year-by-year)*

From 2014 to 2023, we calculate the Log Winsorized Total Asset value for each company annually. Based on these values, companies are classified as either "Big" or "Mid-sized". With the year-by-year classification, companies are divided into two groups based on their Log Winsorized Total Asset values. If a company falls in the middle, it joins the group with the closest value.

Pros	Cons
• Captures annual fluctuations, giving a dynamic picture.	• Some companies shift between categories from year to year and lack consistency.
• Reflects market and economic changes each year.	• The changing classification makes it harder to compare long-term trends.
• Provides detailed yearly insights.	• More data processing/ validation needed
• Tracks a company's growth or decline more precisely.	

To illustrate distribution and normality, this approach is complemented with histograms & QQ plots per year of the values (R Core Team, 2021; The jamovi project, 2022):

Figure F1: Histogram of Log Total Assets Year-by-year method

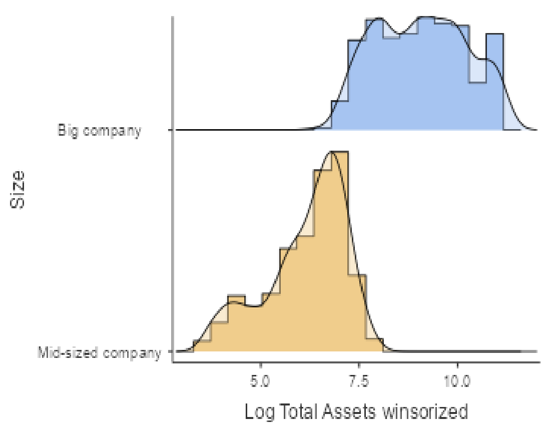
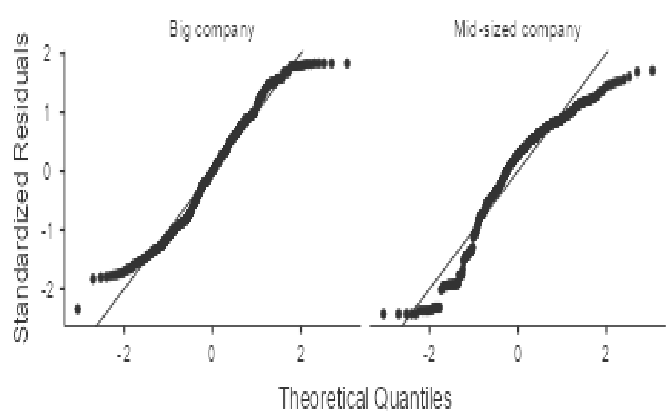


Figure F2: QQ plot of Year-by-year method



Section F2: *Average classification (Years average)*

The Log Winsorized Total Assets are averaged for each company from 2014 to 2023. Then companies are ranked based on these averages, with the 46th company (the median of the rank list) marking the limit between "Big" and "Mid-sized.", joins the group with the closest value. With this classification of a ten-year average, each company gets a single classification over the entire period.

Pros

- Provides a single classification.
- Long-term trend analysis is more straightforward to be done.
- Reduces the complexity of managing data

Cons

- Smooth out the annual variability and hide short-term fluctuations.
- Key annual shifts are missed as company size changes.
- Over time, this blurs the lines between "Big" and "Mid-sized."
- Results in uneven distributions in histograms and QQ plots.

Figure F3: Histogram of Log Total Assets Years average method

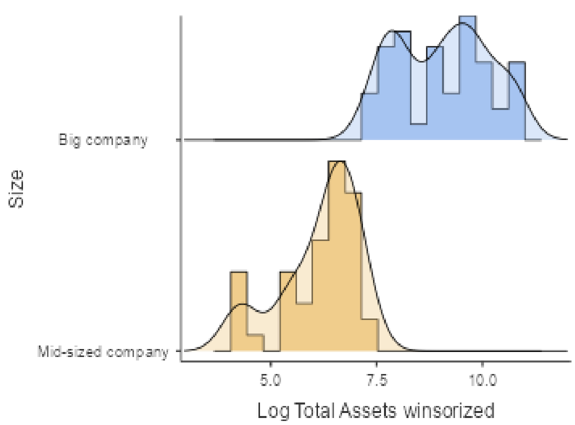
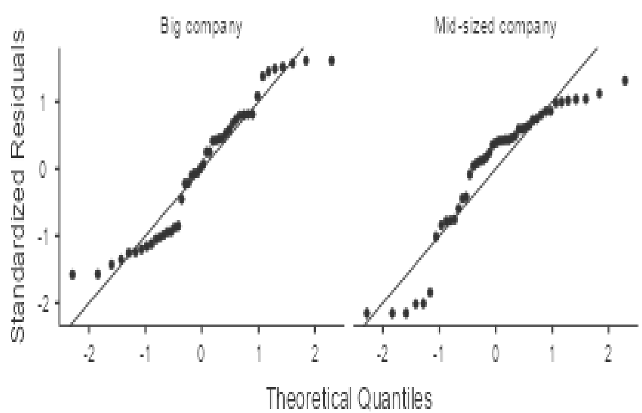


Figure F4: QQ plot of Years average method



***Section F3**: Conclusion on Size*

The year-by-year method was selected due to its yearly detail. Despite potential inconsistency with respect to companies being assigned to different groups across the years, it provides a good overview of what occurs over time. This option is supported by our analysis - histograms (Figure F1 and Figure F3) and QQ plots (Figure F2 and Figure F4). When data distribution is not uniform, QQ plot points tend to depart from the reference line (i.e., non-normality). The average data shows extreme values / skewness causing the QQ plot to have more diverging points.

Appendix G: Initial and Final Sample

Table G1: Initial Company Sample and Reasons for Exclusion (Companies Marketcap, 2024)

Company	*Note*
NVIDIA	Included
TSMC	Included
Broadcom	Included
Samsung	Included
ASML	Included
AMD	Included: The negative equity occurred only in one year (2015), & AMD is a major player.
Intel	Included
QUALCOMM	Included
Applied Materials	Included
Texas Instruments	Included
Arm Holdings	Excluded: 2+ years missing data points
Lam Research	Included
Tokyo Electron	Included
Analog Devices	Included
Micron Technology	Included
KLA	Included
Synopsys	Included
SK Hynix	Included
NXP Semiconductors	Included
Marvell Technology Group	Included
MediaTek	Included
Infineon	Included
Microchip Technology	Included
STMicroelectronics	Included
Monolithic Power Systems	Included
ON Semiconductor	Included
Advantest	Included
Disco Corp.	Included
ASM International	Included
GlobalFoundries	Excluded: 2+ years missing data points
Renesas Electronics	Included
SMIC	Included
Lasertec	Included
ASE Group	Excluded: 2+ years missing data points
Entegris	Included
United Microelectronics	Included
Skyworks Solutions	Included
BE Semiconductor	Included
Novatek Microelectronics	Included
Qorvo	Included
Advanced Micro-Fabrication Equipment	Excluded: 2+ years missing data points
Lattice Semiconductor	Included
Alchip Technologies	Included
Onto Innovation	Included
Realtek	Included
Coherent	Included
GlobalWafers	Included

(to be continued)

Table G1 *(continued)*

Company	Note
Amkor Technology	Excluded: Notable errors in six out of ten financial statements
Rohm	Included
eMemory Technology	Included
Nanya Technology	Excluded: 2+ years missing data points
Global Unichip Corp.	Included
MACOM	Included
Allegro MicroSystems	Excluded: 2+ years missing data points
Technoprobe	Excluded: 2+ years missing data points
Soitec	Included: The negative value was minimal (-9,000) and confined to one year (2015)
Silergy	Included
Cirrus Logic	Included
ASM Pacific Technology	Included
Nova Measuring Instruments	Excluded: Headquarters region falls outside the research scope
Silicon Labs	Included
Power Integrations	Included
Aixtron	Included
Vanguard International Semiconductor	Included
Tower Semiconductor	Excluded: Headquarters region falls outside the research scope
Axcelis Technologies	Included
Credo Technology	Excluded: Headquarters region falls outside the research scope
Powerchip Semiconductor Manufacturing	Excluded: 2+ years missing data points
Melexis NV	Included
Camtek	Excluded: headquarters region falls outside the research scope
Sino-American Silicon Products	Included
Wolfspeed	Included
FormFactor	Included
VisEra Technologies	Excluded: 2+ years missing data points
Vishay Intertechnology	Included
Siltronic	Included: Only 1 year of missing data
Kulicke and Soffa Industries	Included
Maruwa	Included
SiTime	Excluded: 2+ years missing data points
Silicon Motion	Included
Ambarella	Included
WIN Semiconductors	Included
Veeco	Included
Formosa Sumco Technology	Included
Ultra Clean Holdings Inc	Included
Photronics	Included
Nordic Semiconductor	Included
Cohu	Included
PDF Solutions	Included
Semtech	Included
Ichor Systems	Excluded: 2+ years missing data points
ACM Research	Included: Only 1 year of missing data
indie Semiconductor	Excluded: 2+ years missing data points
X-FAB	Included
Navitas Semiconductor	Excluded: 2+ years missing data points
SMART Global Holdings	Included: The negative value was trivial (-1.000) and occurred only in one year (2016).

(to be continued)

Table G1 *(continued)*

Company	Note
ChipMOS Technologies	Included
Himax	Included
AT & S	Included
u-blox	Included
nLIGHT	Included: Only 1 year of missing data
Alpha & Omega Semiconductor	Included
CEVA	Included
Aehr Test Systems	Included: The negative value was insignificant (-1,000) and limited to one year (2015).
Xperi	Excluded: 2+ years missing data points
SkyWater Technology	Excluded: 2+ years missing data points
BrainChip	Excluded: Headquarters region falls outside the research scope
Transphorm	Excluded: 2+ years missing data points
Valens Semiconductor	Excluded: Headquarters region falls outside the research scope
Arteris	Excluded: 2+ years missing data points
Magnachip	Excluded: Exhibited negative equity for 6 consecutive years from 2014 to 2019.
Everspin Technologies	Included: The negative value was minor (-6,000) and restricted to one year (2015).
Atomera	Excluded: Reported negative equity for 2 fiscal years, 2014 and 2015.
Quicklogic	Included
Sequans Communications	Excluded: Faced negative equity for three fiscal years from 2019 to 2021.
Pixelworks	Included
INTEST Corp	Included
AXT Inc	Included
POET Technologies	Excluded: Headquarters region falls outside the research scope

Table G2: Further exclusion on singular statements

Company	*Year*	*Note*
Taiwan Semiconductor Manufacturing	2023	Difference in reported and calculated Non-current Assets
SK Hynix Inc	2023	Difference in reported and calculated Non-current Assets and Non-current Liabilities
Infineon Technologies AG	2017	Difference in reported and calculated Non-current Liabilities
Infineon Technologies AG	2016	Difference in reported and calculated Non-current Liabilities
STMicroelectronics NV	2014	Difference between Total Assets and Total Liabilities reported
Siltronic AG	2014	Difference in reported and calculated Non-current Liabilities
Soitec SA	2015	Difference in reported and calculated Non-current Liabilities
ACM Research Inc	2023	Difference in reported and calculated Non-current Assets
BE Semiconductor Industries NV	2015	Difference in reported and calculated Total Assets
Aixtron SE	2023	Difference in reported and calculated Current and Non-current Assets
Aixtron SE	2014	Difference in reported and calculated Non-current Assets
nLIGHT Inc	2015	Omission of data
nLIGHT Inc	2014	Omission of data
PDF Solutions Inc	2023	Difference in reported and calculated Total Assets
Everspin Technologies Inc	2023	Difference in reported and calculated Current and Non-current Assets

Table G3: Final Sample

Company	*Region*	*Sector*	*Size*
Samsung Electronics Co Ltd (005930.KS)	East Asia	IDM	Big company
Intel Corp (INTC.O)	USA	IDM	Big company
Taiwan Semiconductor Manufacturing Co Ltd (2330.TW)	East Asia	Foundry	Big company
SK Hynix Inc (000660.KS)	East Asia	IDM	Big company
Micron Technology Inc (MU.O)	USA	IDM	Big company
Broadcom Inc (AVGO.O)	USA	Fabless	Big company
Qualcomm Inc (QCOM.O)	USA	Fabless	Big company
ASML Holding NV (ASML.AS)	Europe	SME&S	Big company
NXP Semiconductors NV (NXPI.O)	Europe	IDM	Big company
Texas Instruments Inc (TXN.O)	USA	IDM	Big company
Analog Devices Inc (ADI.O)	USA	IDM	Big company
Applied Materials Inc (AMAT.O)	USA	SME&S	Big company
NVIDIA Corp (NVDA.O)	USA	Fabless	Big company
Infineon Technologies AG (IFXGn.DE)	Europe	IDM	Big company
Semiconductor Manufacturing International Corp (0981.HK)	East Asia	Foundry	Big company
MediaTek Inc (2454.TW)	East Asia	Fabless	Big company
STMicroelectronics NV (STMMI.MI)	Europe	IDM	Big company
Lam Research Corp (LRCX.O)	USA	SME&S	Big company
United Microelectronics Corp (2303.TW)	East Asia	Foundry	Big company
Renesas Electronics Corp (6723.T)	East Asia	IDM	Big company
Tokyo Electron Ltd (8035.T)	East Asia	Foundry	Big company
Microchip Technology Inc (MCHP.O)	USA	IDM	Big company
Marvell Technology Inc (MRVL.OQ)	USA	Fabless	Big company
Advanced Micro Devices Inc (AMD.O)	USA	Fabless	Big company
Rohm Co Ltd (6963.T)	East Asia	IDM	Big company
KLA Corp (KLAC.O)	USA	Foundry	Big company
ON Semiconductor Corp (ON.O)	USA	IDM	Big company
Synopsys Inc (SNPS.O)	USA	Fabless	Big company
Qorvo Inc (QRVO.O)	USA	IDM	Big company
Skyworks Solutions Inc (SWKS.O)	USA	IDM	Big company
Vishay Intertechnology Inc (VSH)	USA	Foundry	Big company
Wolfspeed Inc (WOLF.K)	USA	IDM	Big company
Sino-American Silicon Products Inc (5483.TWO)	East Asia	Foundry	Big company
Coherent Corp (COHR.K)	USA	IDM	Big company
Entegris Inc (ENTG.O)	USA	SME&S	Big company
Advantest Corp (6857.T)	East Asia	SME&S	Big company
ASM International NV (ASMI.AS)	Europe	SME&S	Big company
ASMPT Ltd (0522.HK)	East Asia	SME&S	Big company
Disco Corp (6146.T)	East Asia	SME&S	Big company
Globalwafers Co Ltd (6488.TWO)	East Asia	SME&S	Big company
Realtek Semiconductor Corp (2379.TW)	East Asia	Fabless	Big company
AT & S Austria Technologie & Systemtechnik AG (ATSV.VI)	Europe	SME&S	Big company
Siltronic AG (WAFGn.DE)	Europe	SME&S	Big company
Novatek Microelectronics Corp (3034.TW)	East Asia	Fabless	Big company
Vanguard International Semiconductor Corp (5347.TWO)	East Asia	Foundry	Big company
Silicon Laboratories Inc (SLAB.O)	USA	Fabless	Big company
Cirrus Logic Inc (CRUS.O)	USA	Fabless	Mid-sized company

(to be continued)

Table G3 *(continued)*

Company	Region	Sector	Size
WIN Semiconductors Corp (3105.TWO)	East Asia	Foundry	Mid-sized company
ChipMOS Technologies Inc (8150.TW)	East Asia	IDM	Mid-sized company
MACOM Technology Solutions Holdings Inc (MTSI.O)	USA	IDM	Mid-sized company
Kulicke and Soffa Industries Inc (KLIC.O)	East Asia	SME&S	Mid-sized company
Photronics Inc (PLAB.O)	USA	SME&S	Mid-sized company
Semtech Corp (SMTC.O)	USA	Fabless	Mid-sized company
Himax Technologies Inc (HIMX.O)	East Asia	Fabless	Mid-sized company
Soitec SA (SOIT.PA)	Europe	SME&S	Mid-sized company
Veeco Instruments Inc (VECO.O)	USA	SME&S	Mid-sized company
Monolithic Power Systems Inc (MPWR.O)	USA	Fabless	Mid-sized company
BE Semiconductor Industries NV (BESI.AS)	Europe	SME&S	Mid-sized company
Ultra Clean Holdings Inc (UCTT.O)	USA	SME&S	Mid-sized company
X Fab Silicon Foundries EV (XFAB.PA)	Europe	Foundry	Mid-sized company
Formosa Sumco Technology Corp (3532.TW)	East Asia	SME&S	Mid-sized company
SMART Global Holdings Inc. (SGH.O)	USA	Fabless	Mid-sized company
Cohu Inc (COHU.O)	USA	SME&S	Mid-sized company
FormFactor Inc (FORM.O)	USA	SME&S	Mid-sized company
Lattice Semiconductor Corp (LSCC.O)	USA	Fabless	Mid-sized company
Power Integrations Inc (POWI.O)	USA	Fabless	Mid-sized company
Silicon Motion Technology Corp (SIMO.O)	East Asia	Fabless	Mid-sized company
Onto Innovation Inc (ONTO.K)	USA	SME&S	Mid-sized company
Alpha and Omega Semiconductor Ltd (AOSL.O)	USA	IDM	Mid-sized company
Aixtron SE (AIXGn.DE)	Europe	SME&S	Mid-sized company
MARUWA Co Ltd (5344.T)	East Asia	SME&S	Mid-sized company
Lasertec Corp (6920.T)	East Asia	SME&S	Mid-sized company
Axcelis Technologies Inc (ACLS.O)	USA	SME&S	Mid-sized company
Ambarella Inc (AMBA.O)	USA	Fabless	Mid-sized company
U Blox Holding AG (UBXN.S)	Europe	Fabless	Mid-sized company
Melexis NV (MLXS.BR)	Europe	Fabless	Mid-sized company
Silergy Corp (6415.TW)	East Asia	Fabless	Mid-sized company
Nordic Semiconductor ASA (NOD.OL)	Europe	Fabless	Mid-sized company
Global Unichip Corp (3443.TW)	East Asia	Fabless	Mid-sized company
CEVA Inc (CEVA.O)	USA	Fabless	Mid-sized company
AXT Inc (AXTI.O)	USA	SME&S	Mid-sized company
PDF Solutions Inc (PDFS.O)	USA	SME&S	Mid-sized company
ACM Research Inc (ACMR.O)	USA	SME&S	Mid-sized company
nLIGHT Inc (LASR.O)	USA	SME&S	Mid-sized company
Alchip Technologies Ltd (3661.TW)	East Asia	Fabless	Mid-sized company
eMemory Technology Inc (3529.TWO)	East Asia	Fabless	Mid-sized company
Pixelworks Inc (PXLW.O)	USA	Fabless	Mid-sized company
Quicklogic Corp (QUIK.O)	USA	Fabless	Mid-sized company
Everspin Technologies Inc (MRAM.O)	USA	IDM	Mid-sized company
inTest Corp (INTT.K)	USA	SME&S	Mid-sized company
Aehr Test Systems (AEHR.O)	USA	SME&S	Mid-sized company

Table G4: Sample size per region and year

Region	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	Total Obs.
USA	46	47	48	48	48	48	48	48	48	34	463
Europe	11	12	13	13	14	14	14	14	14	9	128
East Asia	29	28	29	29	29	29	29	29	29	12	272
Total Obs.	86	87	90	90	91	91	91	91	91	55	863

Table G5: Sample size per sector and year

Sector	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	Total Obs.
IDM	20	20	20	20	21	21	21	21	21	13	198
Foundry	10	10	10	10	10	10	10	10	10	4	94
Fabless	29	29	29	29	29	29	29	29	29	21	282
SME&S	27	28	31	31	31	31	31	31	31	17	289
Total Obs.	86	87	90	90	91	91	91	91	91	55	863

Table G6: Sample size per firm size and year

Firm Size	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	Total Obs.
Big company	43	43	45	45	46	46	46	46	45	28	433
Mid-sized company	43	44	45	45	45	45	45	45	46	27	430
Total Obs.	86	87	90	90	91	91	91	91	91	55	863

Appendix H: Descriptive statistics

Table H1: Detailed listing and explanation of variables

Variable	*Acronym*	*Categories and Measurements*
General firm characteristics (Categorical variables)		
Region	REGION	USA, East Asia or Europe
SC industry sector	SECTOR	IDM, Fabless, Foundry or SME&S
Firm size	FMSIZE	Big companies or Mid-sized companies
Fiscal year	YEAR	2014, 2015, 2016, 2018, 2019, 2020, 2021, 2022, 2023
Financial characteristics (Continuous variables)		
Gross Margin	GROSS	$\frac{Gross\ Profit\ (Industrials\ and\ Property;\ Total)}{Revenue\ from\ Business\ Activities}$
Current Asset Turnover	CAT	$\frac{Revenue\ from\ Business\ Activities}{Current\ Assets}$
R&D ratio	R&Dg	$\frac{Research\ \&\ Development\ Expense\ (Expensed\ \&\ Capitalized;Total)}{Gross\ Profit\ (Industrials\ and\ Property;\ Total)}$
SG&A minus R&D ratio	SG&A*	$\frac{Selling, General\ \&\ Administrative\ Expenses\ -\ Research\ \&\ Development}{Gross\ Profit\ (Industrials\ and\ Property;\ Total)}$
Property, Plant and Equipment ratio	PPEa	$\frac{Property, Plant\ \&\ Equipment\ (Net;\ Total)}{Total\ Assets}$
Intangible Assets ratio	INTGa	$\frac{Intangible\ Assets\ (Net;\ Total)}{Total\ Assets}$

Table H2: Descriptive statistics of the Income statement dataset

Variable	*Mean*	*Median*	*Range*	*Standard Deviation*	*1st Quartile*	*2nd Quartile*	*3rd Quartile*
Revenue from Business Activities - Total	1.00	1.00	0.00	0.00	1.00	1.00	1.00
Cost of Operating Revenue	0.57	0.57	1.23	0.15	0.48	0.57	0.66
Gross Profit - Industrials/Property - Total	0.43	0.43	1.23	0.15	0.34	0.43	0.52
Research & Development Expense - Expensed & Capitalized - Total	0.16	0.13	1.20	0.13	0.08	0.13	0.20
Selling, General & Administrative Expenses minus R&D expenses	*0.15*	*0.12*	*0.91*	*0.10*	*0.08*	*0.12*	*0.19*
Other Operating Expense/(Income) - Net	0.00	0.00	0.27	0.01	0.00	0.00	0.00
Operating Profit before Non-Recurring Income/Expense (EBIT)	0.13	0.14	2.22	0.20	0.07	0.14	0.24
Non-Operating Income/(Expenses) - Total	-0.01	0.00	1.25	0.06	-0.02	0.00	0.01
Income before Taxes	0.12	0.13	2.31	0.22	0.05	0.13	0.24
Income Taxes	0.02	0.02	0.54	0.04	0.01	0.02	0.04
Net Income after Tax	0.10	0.12	2.33	0.21	0.04	0.12	0.21
Depreciation, Depletion & Amortization	0.09	0.06	0.50	0.07	0.04	0.06	0.11

Table H3: Descriptive statistics of the Balance sheet dataset

Variable		*Mean*	*Median*	*Range*	*Standard Deviation*	*1st Quartile*	*2nd Quartile*	*3rd Quartile*
Current Assets	Cash & Short-Term Investments	0.31	0.28	0.79	0.16	0.18	0.28	0.41
	Financial Assets - Short-Term	0.01	0.00	0.56	0.05	0.00	0.00	0.00
	Derivative Financial Instruments - Hedging - Short-Term	0.00	0.00	0.05	0.00	0.00	0.00	0.00
	Loans & Receivables - Net - Short-Term	0.12	0.10	0.42	0.06	0.07	0.10	0.14
	Inventories - Total	0.12	0.11	0.70	0.08	0.07	0.11	0.16
	Prepaid Expenses - Short-Term	0.01	0.01	0.13	0.01	0.00	0.01	0.02
	Assets Held for Sale/Discontinued Operations - Short-Term	0.00	0.00	0.10	0.01	0.00	0.00	0.00
	Other Current Assets - Total	0.01	0.00	0.33	0.03	0.00	0.00	0.01
	Total Current Assets	***0.57***	***0.56***	***0.88***	***0.19***	***0.43***	***0.56***	***0.72***
Non-Current Assets	Investments - Long-Term	0.02	0.00	0.30	0.04	0.00	0.00	0.02
	Investments in Associates, Joint Ventures and Unconsolidated Subsidiaries	0.01	0.00	0.58	0.05	0.00	0.00	0.00

(to be continued)

Table H3 *(continued)*

Variable	Mean	Median	Range	Standard Deviation	1st Quartile	2nd Quartile	3rd Quartile	Variable
Non-Current Assets	Receivables & Loans - Long-Term	0.00	0.00	0.09	0.01	0.00	0.00	0.00
	Derivative Financial Instruments - Hedging - Long-Term	0.00	0.00	0.01	0.00	0.00	0.00	0.00
	Property, Plant & Equipment - Net - Total	0.20	0.14	0.79	0.16	0.08	0.14	0.29
	Assets Held for Sale/Discontinued Operations - Long-Term	0.00	0.00	0.13	0.01	0.00	0.00	0.00
	Other Non-Current Assets - Total	0.03	0.02	0.25	0.03	0.01	0.02	0.04
	Intangible Assets - Total - Net	0.17	0.08	0.83	0.19	0.02	0.08	0.26
	Total Non-Current Assets minus Property, Plant & Equipment and minus Intangible Assets	*0.07*	*0.05*	*0.58*	*0.07*	*0.02*	*0.05*	*0.09*
	Total Non-Current Assets	***0.43***	***0.44***	***0.88***	***0.19***	***0.28***	***0.44***	***0.57***
	Total Assets	***1.00***	***1.00***	***0.00***	***0.00***	***1.00***	***1.00***	***1.00***
Current Liabilities	Trade Accounts Payable & Accruals - Short-Term	0.10	0.09	0.52	0.06	0.07	0.09	0.12
	Short-Term Debt & Notes Payable	0.02	0.00	0.52	0.06	0.00	0.00	0.00
	Current Portion of Long-Term Debt including Capitalized Leases	0.02	0.00	0.28	0.03	0.00	0.00	0.02
	Derivative Liabilities - Hedging - Short-Term	0.00	0.00	0.04	0.00	0.00	0.00	0.00
	Liabilities Held for Sale/Discontinued Operations - Short-Term	0.00	0.00	0.05	0.00	0.00	0.00	0.00
	Income Taxes - Payable - Short-Term	0.01	0.00	0.09	0.01	0.00	0.00	0.01
	Dividends/Distributions Payable	0.00	0.00	0.06	0.01	0.00	0.00	0.00
	Operating Lease Liabilities - Current Portion/Short-Term	0.00	0.00	0.05	0.00	0.00	0.00	0.00
	Other Current Liabilities - Total	0.05	0.03	0.56	0.06	0.01	0.03	0.07
	Total Current Liabilities	***0.19***	***0.16***	***0.81***	***0.11***	***0.12***	***0.16***	***0.23***

(to be continued)

Table H3 *(continued)*

Variable	Mean	Median	Range	Standard Deviation	1st Quartile	2nd Quartile		3rd Quartile
Non-Current Liabilities	Accounts Payable including Accrued Expenses - Long-Term	0.00	0.00	0.06	0.01	0.00	0.00	0.00
	Debt - Long-Term - Total	0.13	0.06	0.66	0.15	0.00	0.06	0.22
	Liabilities Held for Sale/Discontinued Operations - Long-Term	0.00	0.00	0.04	0.00	0.00	0.00	0.00
	Derivative Liabilities - Hedging - Long-Term	0.00	0.00	0.05	0.00	0.00	0.00	0.00
	Deferred Tax & Investment Tax Credits - Long-Term	0.01	0.00	0.14	0.02	0.00	0.00	0.01
	Operating Lease Liabilities - Long-Term	0.00	0.00	0.10	0.01	0.00	0.00	0.00
	Other Non-Current Liabilities - Total	0.04	0.02	0.57	0.05	0.01	0.02	0.05
	Minority Interest - Non-Equity	0.00	0.00	0.28	0.01	0.00	0.00	0.00
	Total Non-Current Liabilities	***0.18***	***0.13***	***0.71***	***0.16***	***0.04***	***0.13***	***0.28***
	Total Liabilities	***0.37***	***0.36***	***1.47***	***0.19***	***0.22***	***0.36***	***0.49***
Shareholders' Equity	Shareholders' Equity - Attributable to Parent Shareholders - Total	0.62	0.63	1.47	0.19	0.49	0.63	0.77
	Minority Interest - Equity	0.01	0.00	0.55	0.04	0.00	0.00	0.00
	Hybrid Financial Instrument - Equity Portion	0.00	0.00	0.11	0.01	0.00	0.00	0.00
	Total Shareholders' Equity - including Minority Interest & Hybrid Debt	***0.63***	***0.64***	***1.47***	***0.19***	***0.51***	***0.64***	***0.78***
	Total Liabilities & Equity	***1.00***	***1.00***	***0.08***	***0.00***	***1.00***	***1.00***	***1.00***

Appendix I: Winsorization[22]

***Section I1:** Missing data*

Missing data may result in loss of efficiency, complications in how the researcher deals with data and potentially introduce bias as well (Kaiser, 2014). To counter this issue, the current process iterates every observation that is represented by a year of financial statement and if there is a missing year, simple methods are used, by identifying the missing data points and filling them with 0 for those values. This is due to the fact that if there is no year presence it means there is a missing financial statement for that year, and to assume the data points using mean or median values, or more advanced techniques based on machine learning algorithms such as *k*-nearest neighbour and neural networks, is not viable (Doshi, 2010; Kaiser, 2014).

***Section I2:** Winsorization*

The next phase uses Winsorizing to reduce the impact of extreme values. Winsorization is a series of modifications to reduce the impact of outliers. It replaces values below and above certain percentiles with the values at those percentiles (Wilcox, 2005).

To winsorize the variable at 80% by year, we adjust the values that fall below the 10th percentile and above the 90th percentile each year. Specifically, any value below the 10th percentile is set to match the value at the 10th percentile. Similarly, values above the 90th percentile are capped at the 90th percentile. This process ensures that extreme values are brought in line with more typical observations, reducing the impact of outliers on the analysis. This method is often applied to make statistical analysis more accurate by ensuring that extreme values are brought in line with more typical observations. It is extremely useful particularly when dealing with abnormal distributions or data with multiple outliers, such as ours (Fernàndez et al., 2002).

***Section I3:** Additional adjustment:*

We proceed with adjusting the remaining outliers after winsorization to fall within a calculated yearly range. The winsorized dataset is processed to take outliers that still fall outside a reasonable range and replace them with the highest/lowest non-outlier value, hence the value at the upper/lower quartile plus/minus 1.5 times the interquartile range.

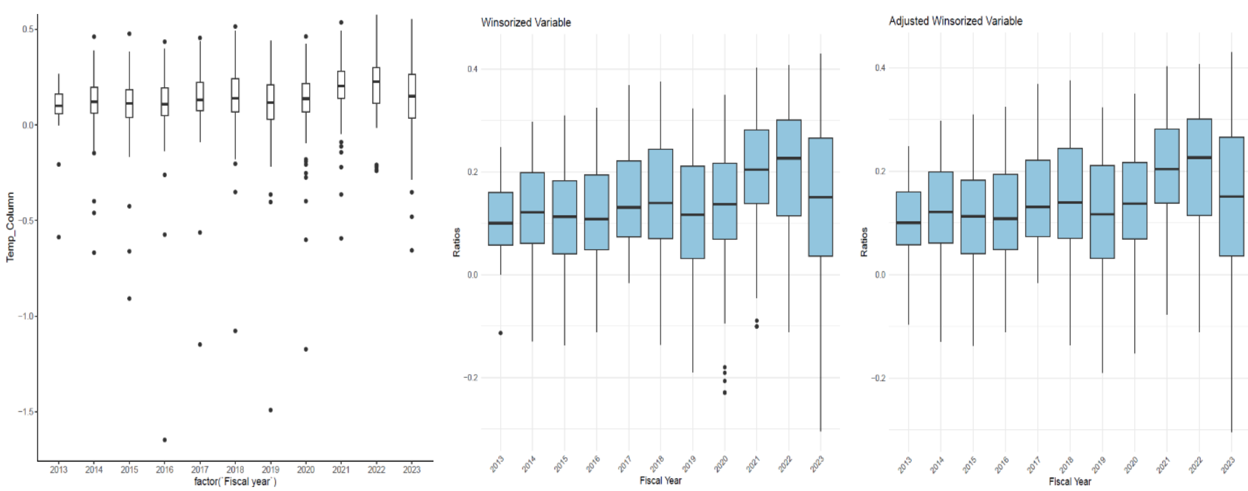
[22] All of the following operations were executed via R script *(Appx. Z; Code 8)*.

Section I4: *Example of how the results might look for an EBIT-to-revenue ratio*

Table I1: Winsorization and Adjustment (E.g. Ambarella Inc (AMBA.O)) that is repeated with all the EBIT/Revenue ratios

Company	*Fiscal year*	**Original _EBIT**	**Winsorized_ variable**	*Winsorize _Made*	**Adjusted_Win _variable**	*Adjustment _Made*
…	…	…	…	…	…	…
Ambarella Inc (AMBA.O)	*2023*	**-0.656**	**-0.305**	*Yes*	**-0.305**	*No*
Ambarella Inc (AMBA.O)	*2022*	**-0.220**	**-0.111**	*Yes*	**-0.111**	*No*
Ambarella Inc (AMBA.O)	*2021*	**-0.089**	**-0.089**	*No*	**-0.077**	*Yes*
Ambarella Inc (AMBA.O)	*2020*	**-0.274**	**-0.229**	*Yes*	**-0.152**	*Yes*
Ambarella Inc (AMBA.O)	*2019*	**-0.217**	**-0.189**	*Yes*	**-0.189**	*No*
Ambarella Inc (AMBA.O)	*2018*	**-0.177**	**-0.137**	*Yes*	**-0.137**	*No*
Ambarella Inc (AMBA.O)	*2017*	**0.083**	**0.083**	*No*	**0.083**	*No*
Ambarella Inc (AMBA.O)	*2016*	**0.195**	**0.195**	*No*	**0.195**	*No*
Ambarella Inc (AMBA.O)	*2015*	**0.268**	**0.268**	*No*	**0.268**	*No*
Ambarella Inc (AMBA.O)	*2014*	**0.238**	**0.238**	*No*	**0.238**	*No*
…	…	…	…	…	…	…

Figure I1-I3: Original values, Winsorization by year and Adjustment



Appendix J: χ^2 test of independence/association

Section J1: *Theoretical framework: χ^2 test of independence/association*

The χ^2 test is appropriate in this MT in order to analyze the relationships between discrete variables like FMSIZE, SECTOR, REGION, and YEAR, where each variable represents a category rather than a continuous measurement. The steps of statistical hypothesis testing are as follows (Msuha, 2019):

- the main hypothesis H_0 and alternative hypothesis H_1 are formulated;
- the statistical criterion is selected, by means of which the hypothesis will be tested;
- the value of the significance level α is set;
- the boundaries of the hypothesis acceptance area are found;
- the conclusion about acceptance or rejection of the main hypothesis H_0 is made.

The null hypothesis states that the difference between the empirical and hypothetical rules (e.g., normal) is significant and, therefore, the considered random variable cannot be considered normally distributed with a high probability (Msuha, 2019). The alternative hypothesis assumes the absence of significant differences, and consequently, the presence of consistency between the empirical and hypothetical distributions. Thus, the criterion of agreement, like other statistical criteria, should confirm or reject the null hypothesis(Msuha, 2019). Two statistical hypotheses are proposed: null (H_0: categorical variables A and B are independent) and alternative hypotheses (H_1) (H_1: categorical variables A and B are related to each other). The following hypotheses are set:

H_1 (a): *There is an association between FMSIZE and at least one of the following variables: (i) YEAR, (ii) SECTOR, or (iii) REGION.*

H_1 (b): *There is an association between SECTOR and at least one of the following variables: (i) YEAR, (ii) FMSIZE, or (iii) REGION.*

H_1 (c): *There is an association between a YEAR and at least one of the following variables: (i) SECTOR, (ii) FMSIZE, or (iii) REGION.*

H_1 (d): *There is an association between a REGION and at least one of the following variables: (i) SECTOR, (ii) FMSIZE, or (iii) YEAR.*

The tests' validity depends on the number of observations in the sample and expected frequencies, with guidelines suggesting a minimum sample SIZE of 40 for 2x2 tables (Pandis, 2016). At first χ^2 the test was to be conducted directly on the number of companies, but upon examination, it was evident that the number of observations in the sample was insufficiently representative to perform the test effectively (Pandis, 2016). *(see Table J1).*

Table J1: Frequencies of companies by REGION and SECTOR

	IDM	Fabless	Foundry	SME&S	**Total No. of Companies**
USA	*13*	*17*	*2*	*16*	*48*
East Asia	*5*	*9*	*7*	*8*	*29*
Europe	*3*	*3*	*1*	*7*	*14*
Total No. of Companies	*21*	*29*	*10*	*31*	*91*

Hence, we proceeded by considering the number of financial statements in the sample. Each company provides between 8-10 statements. Thus, this can be a sufficient representation of the number of companies for each categorical variable *(see Table J2)*.

Table J2: Frequencies of observations by REGION and SECTOR

	IDM	Fabless	Foundry	SME&S	**Total Obs.**
USA	*124*	*167*	*19*	*153*	*463*
East Asia	*47*	*86*	*65*	*74*	*272*
Europe	*27*	*29*	*10*	*62*	*128*
Total Obs.	*198*	*282*	*94*	*289*	*863*

According to the probability multiplication theorem, when two random variables A and B are independent, the probability of getting a joint event is P(AB) = P(A) × P(B). Thus, expected values can be calculated using the probability multiplication formula and summarized in Table J3:

Table J3: Expected Frequencies of Observations by REGION and SECTOR

	IDM	Fabless	Foundry	SME&S		**Total Obs.**
USA	*106.23*	*151.29*	*50.43*	155.05	$= \frac{198}{863} \times \frac{463}{863} \times 863$	*463*
East Asia	*62.41*	*88.88*	*29.63*	*91.09*		*272*
Europe	*29.37*	*41.83*	*13.94*	*42.86*		*128*
Total Obs.	*198*	*282*	*94*	*289*		*863*

Using this table, we can calculate Pearson's χ^2 statistics (Plackett, 1983) using the [J1] formula to achieve the results in Table J4:

$$\chi^2 = \sum \frac{E_{ij}(O_{ij} - E_{ij})^2}{E_{ij}} \qquad [J1]$$

Table J4: Sum of Squares by REGION and SECTOR

		IDM	Fabless	Foundry
USA	*2.9736*	*1.6306*	*19.5893*	SME&S
East Asia	*3.8030*	*0.0934*	*42.2339*	*0.0271*
Europe	*0.1908*	*3.9332*	*1.1146*	*3.2053*

$\boldsymbol{\chi^2 = 87.3373783454456}$

Table J5: χ^2 tests Results

Variables	*χ2*	*p-value*	*Interpretation*
FMSIZE - YEAR	0.07	1.0E+00	There is no significant association at α=0.05 between company FMSIZE and YEAR.
FMSIZE - SECTOR	156.11	1.3E-33	There is significant association at α=0.05 between company FMSIZE and the SECTOR.
FMSIZE - REGION	6.62	3.6E-02	There is significant association at α=0.05 between company FMSIZE and the REGION.
YEAR - SECTOR	1.82	1.0E+00	There is no significant association at α=0.05 between YEAR and the SECTOR.
YEAR - REGION	3.10	1.0E+00	There is no significant association at α=0.05 between YEAR and REGION.
REGION - SECTOR	87.34	1.1E-16	There is significant association at α=0.05 between the SECTOR and the REGION.

The significance level α is the probability of a Type I error. The value of the significance level is usually quite small and is set by the analyst testing the hypothesis. It most often takes values of 0.01 (1%), 0.05 (5%), and 0.1 (10%).

When testing a hypothesis, there is always the possibility that a wrong conclusion will be drawn. There are two types of errors (Banerjee et al., 2009):

- The Type I error of error is the rejection of the main hypothesis when it is true.
- The Type II error of error is the acceptance of the main hypothesis when it is false. Related to the value of the significance level is the value of the confidence level *p*.

Confidence level p is the probability of accepting the correct hypothesis. Until it is proven that the main hypothesis H_0 is false, we consider it true. Therefore, the significance level will determine the probability of accepting the main hypothesis. If the significance level α is the probability of rejecting the true hypothesis, then the probability of accepting the true hypothesis is $p = 1 - \alpha$ (Banerjee et al., 2009).

The Type I error is controlled by us - the probability is set of its occurrence. However, the Type II error cannot be controlled - there is always a probability that an incorrect hypothesis may be accepted. Therefore, in order to avoid undesirable consequences of accepting an incorrect hypothesis, in the case of this MT the main hypothesis is formulated in such a way that the risk of accepting an incorrect hypothesis is minimized (Banerjee et al., 2009).

Following the results, we can reject $\boldsymbol{H_1}$ ***(c)***, but not $\boldsymbol{H_1}$ ***(a)***, $\boldsymbol{H_1}$ ***(b)*** and $\boldsymbol{H_1}$ ***(d)***.

Appendix K: Bivariate analysis[23]

Table K1: Mean values Sector

SECTOR	*GROSS*	*CAT*	*R&Da*	*SG&A**	*PPEa*	*INTGa*	*ROA_win*
IDM	0.422	1.520	0.344	0.325	0.270	0.246	0.089
Foundry	0.340	1.170	0.270	0.252	0.357	0.037	0.119
Fabless	0.511	1.290	0.500	0.330	0.102	0.201	0.099
SME&S	0.399	1.110	0.278	0.404	0.193	0.122	0.093

Table K2: Mean values Region

REGION	*GROSS*	*CAT*	*R&Da*	*SG&A**	*PPEa*	*INTGa*	*ROA_win*
East Asia	0.409	1.15	0.295	0.243	0.257	0.053	0.127
USA	0.464	1.31	0.413	0.422	0.157	0.235	0.075
Europe	0.382	1.36	0.339	0.286	0.227	0.163	0.112

Table K3: Mean values Sector and Region

SECTOR	*REGION*	*GROSS*	*CAT*	*R&Da*	*SG&A**	*PPEa*	*INTGa*	*ROA_win*
IDM	East Asia	0.367	1.420	0.260	0.287	0.372	0.100	0.096
	USA	0.449	1.530	0.368	0.355	0.241	0.284	0.086
	Europe	0.394	1.620	0.379	0.254	0.227	0.323	0.095
Foundry	East Asia	0.342	1.110	0.253	0.181	0.396	0.013	0.120
	USA	0.427	1.240	0.169	0.328	0.190	0.137	0.160
	Europe	0.157	1.370	0.573	0.574	0.422	0.010	0.035
Fabless	East Asia	0.453	1.170	0.453	0.192	0.099	0.065	0.139
	USA	0.551	1.290	0.551	0.405	0.096	0.275	0.068
	Europe	0.446	1.690	0.349	0.308	0.148	0.178	0.153
SME&S	East Asia	0.442	0.997	0.172	0.328	0.245	0.044	0.137
	USA	0.386	1.180	0.329	0.506	0.152	0.164	0.064
	Europe	0.383	1.090	0.278	0.242	0.233	0.112	0.112

Table K4: ANOVA and Assumptions checks

F-value	*GROSS*		*CAT*		*R&Da*		*SG&A**		*PPEa*		*INTGa*		*ROA_win*	
REGION	16.70	***	12.25	***	12.70	***	50.53	***	28.63	***	86.91	***	25.53	***
SECTOR	46.10	***	22.94	***	41.63	***	5.46	**	104.02	***	21.68	***	4.94	***
FMSIZE	1.17		24.06	***	8.29	*	85.18	***	1.30	**	65.09	***	48.87	***
Kruskal-Walli's Test	231.30	***	188.60	***	315.47	***	393.73	***	326.63	***	340.34	***	188.73	***
Levene's Test	7.58	***	4.78	***	4.92	***	3.89	***	9.91	***	14.20	***	2.96	***
Shapiro-Wilk Test	0.97	***	0.98	***	0.73	***	0.78	***	0.87	***	0.82	***	0.99	***

*Note: * p < .05, ** p < .01, *** p < .001*

Section K1: ANOVA results and interpretation[24]

The REGION category shows significant effects on all drivers of ROA, with particularly strong influences on SG&A* and INTGa ratio. In a similar way, the SECTOR variable influences significantly all metrics, in particular, PPEa and R&Dg are most affected, indicating sector-specific investment patterns and expenses *(see Table K4)*.

[23] Executed via JAMOVI project, an open statistical software (R Core Team, 2021; The jamovi project, 2022)
[24] Executed via Python 3 script *(see Appx. Script M10)*

FMSIZE is the company size, and the impact of it varies across the drivers. It is not significant for GROSS and does not affect strongly R&Dg and PPEa *(see Table K4)*.

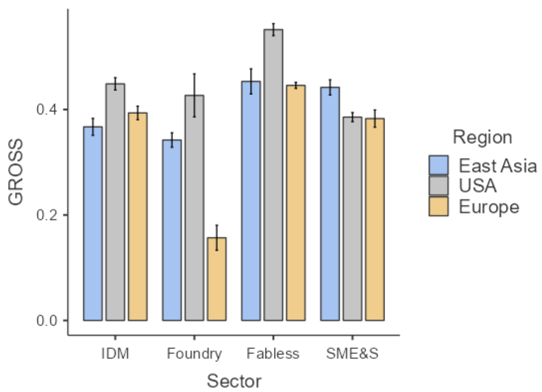
Section K2: *Interpretation of the Tests for Assumptions*[25]

Kruskal-Wallis Test (a non-parametric alternative to ANOVA) (McKight & Najab, 2010) indicates significant differences across all categories, reinforcing the ANOVA results and suggesting that the data does not follow a normal distribution *(see Table K4)*.

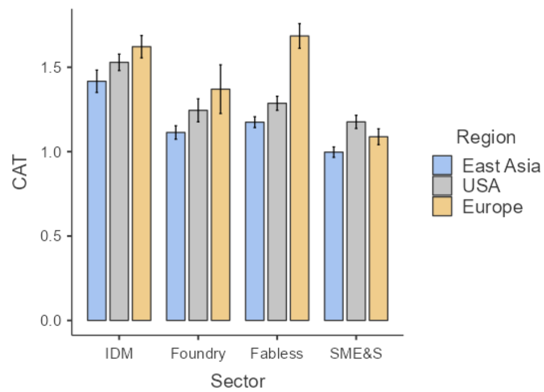
Levene's Test for equality of variances (Nordstokke & Colp, 2014) shows significant results across all tests, suggesting that variance is not consistent across groups, which is important for the validity of ANOVA results *(see Table K4)*.

Shapiro-Wilk Test measures normality (Shapiro & Wilk, 1965), and significant values (less than 0.05) for metrics like R&D and SG&A suggest non-normal distributions, justifying the use of non-parametric tests *(see Table K4)*.

Graph K1: Mean values GROSS

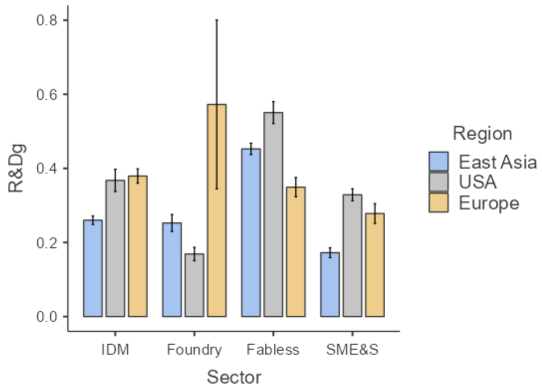


Graph K2: Mean values CAT

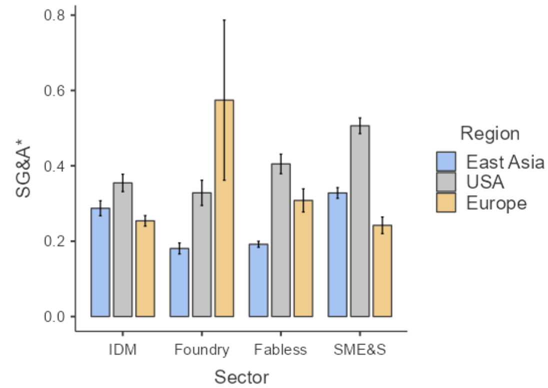


[25] Executed via Python 3 script *(see Appx. Script M11)*

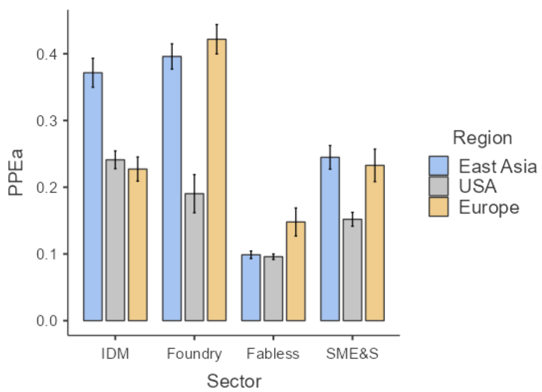
Graph K3: Mean values R&Dg



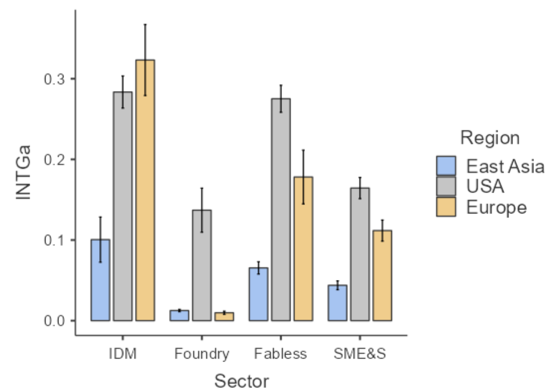
Graph K4: Mean values SG&S*



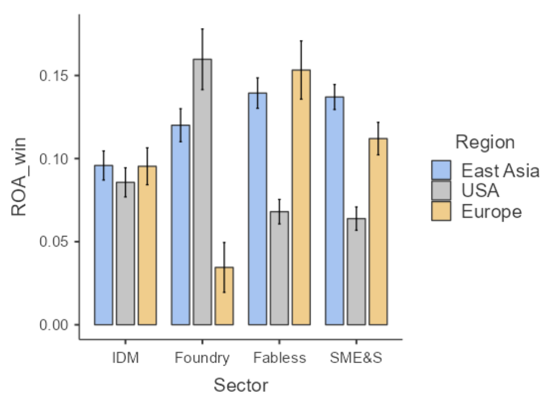
Graph K5: Mean values PPEa



Graph K6: Mean values INTGa



Graph K7: Mean values ROA_win



Appendix L: Multiple linear regression models

Section L1: *Relations of economic variables*

In econometrics, when considering the relationship between two variables X and Y, one of the variables is identified as independent (explanatory) and the other as dependent (explanatory). In this case, a change in the first of them may cause a change in the other. Dependence of this type, expressed by the relation $M(Y|x) = f(x)$, is called regression of Y on X. When considering the dependence of two random variables, we speak of simple regression. The dependence of several variables expressed by a function $M(Y|x_1, x_2, \dots, x_m) = f(x_1, x_2, \dots, x_m)$ is called multiple regression (Lyubimtsev & Lyubimtseva, 2016).

To reflect the fact that the actual values of the dependent variable do not always coincide with its conditional mathematical expectations and may be different for the same value of the explanatory variable (set of explanatory variables), the actual dependence should be supplemented with the summand ϵ, which, is essentially a random variable that indicates the statistical essence of the dependence. It follows that the relationships between the dependent and explanatory variable(s) expressed by the relations $Y = M(Y|x) + \epsilon$ and $Y = M(Y|x_1, x_2, \dots, x_m) + \epsilon$, called regression models (Lyubimtsev & Lyubimtseva, 2016).

Section L2: *Method of Ordinary Least Squares (OLS)*

The relationship presented in equation [L1] is called the theoretical linear regression model; β_0 and β_1 are the theoretical parameters (theoretical regression coefficients; ϵ - random deviation (random error).

To determine the values of theoretical regression coefficients it is necessary to know and use all values of variables X and Y of the general population, which is practically impossible. Thus, it is necessary to be able to estimate the coefficients β_0 and β_1 based on statistical data (sample):

$$M(Y \mid X = x) = \beta_0 + \beta_1 X + \epsilon \qquad [L1]$$

Then based on the sample, we can construct the so-called empirical (sample) regression equation [L2],

$$\hat{y} = b_0 + b_1 \cdot x \qquad [L2]$$

where yˆ is an estimate of the conditional mathematical expectation $M(Y \mid X = x)$; b_0 and b_1 - estimates of the unknown parameters β_0 and β_1, called empirical (sample) regression coefficients. In each particular sample observation, we have $y_i = b_0 + b_1 x_i + e_i$, where deviation e_i is an estimate of the theoretical random deviation ϵ_i (Lyubimtsev & Lyubimtseva, 2016).

In order for a regression analysis based on the OLS to give the best possible results, certain conditions (Gauss-Markov conditions) must be met (Larocca, 2005).

1. The mathematical expectation of the random error in any observation must be equal to zero.
2. The variance of the random error must be constant for all the observations.
3. Random errors must be statistically independent (uncorrelated) among themselves.
4. The explanatory variable x_i is a non-random variable.

Section L3: *Analysis of the quality of the sample equation of multiple linear regression*

R^2: known as the coefficient of determination. It is the proportion of the variance in the response variable that can be explained by the explanatory variables. For multiple regression, the determination coefficient is a non-decreasing function of the number of explanatory variables. Adding a new explanatory variable never reduces the value of R^2. To compensate for such an increase in R^2, an adjusted determination coefficient (R^2_{adj}) is introduced with an adjustment for the number of degrees of freedom (Wooldridge, 2012).

Standard error: The value SE is called the standard error of the regression. This is the average distance by which the observed values deviate from the regression line.

F: The overall F statistic for the regression model. The p-value associated with the overall F statistic tells us whether the regression model as a whole is statistically significant. In other words, it tells us whether the combined two explanatory variables have a statistically significant relationship with the response variable (Wooldridge, 2012).

P-value coefficient. Individual p-values tell us whether each independent variable is statistically significant (Wooldridge, 2012).

Section L4: *Assumptions check*

The Shapiro-Wilk Normality Test (Shapiro & Wilk, 1965) reveals that the data for IDM, SME&S, and Fabless are not normally distributed as indicated by p-values less than 0.001. In contrast, Foundry's data shows a borderline normal distribution with a p-value of 0.021. The consistently low p-values in the Shapiro-Wilk tests indicate potential deviations from normality in the residuals, which could affect the reliability of hypothesis tests and confidence intervals *(see Table L9)*.

The Durbin-Watson Test for Autocorrelation (White, 1992) checks if there are any correlations between residuals in regression models. Values close to 2 suggest no autocorrelation. Here, IDM, Fabless, and SME&S show relatively low autocorrelation, as indicated by their Durbin-Watson statistics being close to 2. However, Foundry exhibits a mild positive autocorrelation (DW statistic of 1.59), indicating some sequential correlation in the residuals *(see Table L10)*.

Finally, the Collinearity Statistics (Studenmund, 2010) reveal the extent of multicollinearity among variables through the Variance Inflation Factor (VIF) and Tolerance. Ideal VIF values are close to 1 (values above 5 or 10 are often cause for concern), and higher Tolerance (close to 1) suggests lower multicollinearity. The data shows that most VIF values across the groups are below critical levels, indicating acceptable levels of multicollinearity, although there are slightly higher VIF values noted particularly for SG&A* and PPEa, which might need further scrutiny *(see Table L11)*.

Section L5: Standardization[26]

The OLS method is also applicable to the multiple regression equation on a standardized scale [L3]:

$$Y^{std} = \beta_0 + \beta_1 X_{i1}^{std} + \beta_2 X_{i2}^{std} + \beta_3 X_{i3}^{std} + \cdots + \beta_m X_{im}^{std} + \epsilon_i \qquad [L3]$$

The Y^{std}, X_{i1}^{std}, X_{i2}^{std}, X_{i3}^{std}, …, X_{im}^{std} are the standardized variables *(see results in Table L4-L8)*: $Y^{std} = \frac{Y-\bar{Y}}{\sigma_Y}$; $X_{im}^{std} = \frac{X_{im}-\bar{X}_{im}}{\sigma_{X_{im}}}$, for which the mean is equal to zero: $\overline{Y^{std}} = \overline{X_{im}^{std}} = 0$, and the mean square deviation is equal to one: $\overline{\sigma_Y} = \overline{\sigma_{X_{im}}} = 1$; β_i is the standardized regression coefficient. The standardized regression coefficients show by how many units the result will change on average if the corresponding factor X_{im} changes by one unit with the average level of other factors unchanged. Because all variables are specified as centred and normalized, standardized regression coefficients β_i can be compared with each other. By comparing them with each other, it is possible to rank the factors according to the strength of their effect on the outcome. This is the main advantage of standardized regression coefficients as opposed to "pure" regression coefficients, which are incomparable in themselves (Gal & Rubinfeld, 2019).

[26] This is a classical method applied in machine learning algorithms to evaluate the relative importance of different predictors within a dataset. By using standardized regression coefficients, analysts can more accurately interpret the impact of each variable, allowing for more informed decisions in feature selection and model optimization (Gal & Rubinfeld, 2019).

Table L1: IDM correlation matrix

Variable	*test*	**GROSS**		**CAT**		**R&Dg**		**SG&A***		**PPEa**		**INTGa**	
GROSS	*Pearson's r*	—											
	Spearman's rho	—											
CAT	*Pearson's r*	0.270	***	—									
	Spearman's rho	0.224	**	—									
R&Dg	*Pearson's r*	0.128		-0.061		—							
	Spearman's rho	-0.222	**	-0.139		—							
SG&A*	*Pearson's r*	-0.128		-0.306	***	0.708	***	—					
	Spearman's rho	-0.276	***	-0.292	***	0.591	***	—					
PPEa	*Pearson's r*	-0.462	***	-0.114		-0.259	***	-0.258	***	—			
	Spearman's rho	-0.477	***	-0.082		-0.282	***	-0.264	***	—			
INTGa	*Pearson's r*	0.339	***	0.466	***	0.034		-0.066		-0.696	***	—	
	Spearman's rho	0.383	***	0.384	***	0.164	*	-0.031		-0.674	***	—	
ROA_win	*Pearson's r*	0.513	***	0.280	***	-0.435	***	-0.645	***	0.145	*	-0.041	
	Spearman's rho	0.526	***	0.333	***	-0.779	***	-0.721	***	0.161	*	0.027	

Note. * p < .05, ** p < .01, *** p < .001

Table L2: Fabless correlation matrix

Variable	*test*	**GROSS**		**CAT**		**R&Dg**		**SG&A***		**PPEa**		**INTGa**	
GROSS	*Pearson's r*	—											
	Spearman's rho	—											
CAT	*Pearson's r*	-0.338	***	—									
	Spearman's rho	-0.324	***	—									
R&Dg	*Pearson's r*	-0.073		-0.31	***	—							
	Spearman's rho	-0.062		-0.239	***	—							
SG&A*	*Pearson's r*	-0.061		-0.257	***	0.708	***	—					
	Spearman's rho	0.082		-0.111		0.192	**	—					
PPEa	*Pearson's r*	0.063		0.198	***	-0.204	***	0.002		—			
	Spearman's rho	-0.022		0.134	*	-0.302	***	0.055		—			
INTGa	*Pearson's r*	0.171	**	0.312	***	-0.023		-0.027		-0.366	***	—	
	Spearman's rho	0.28	***	0.282	***	0.106		0.224	***	-0.36	***	—	
ROA_win	*Pearson's r*	0.166	**	0.258	***	-0.634	***	-0.579	***	0.316	***	-0.259	***
	Spearman's rho	0.087		0.239	***	-0.719	***	-0.684	***	0.249	***	-0.3	***

Note. * p < .05, ** p < .01, *** p < .001

Table L3: Foundries correlation matrix

Variable	*test*	**GROSS**		**CAT**		**R&Dg**		**SG&A***		**PPEa**		**INTGa**	
GROSS	*Pearson's r*	—											
	Spearman's rho	—											
CAT	*Pearson's r*	-0.157		—									
	Spearman's rho	-0.105		—									
R&Dg	*Pearson's r*	-0.416	***	-0.215	*	—							
	Spearman's rho	-0.325	**	-0.336	***	—							
SG&A*	*Pearson's r*	-0.483	***	0.2		0.697	***	—					
	Spearman's rho	-0.603	***	0.288	**	0.171		—					
PPEa	*Pearson's r*	-0.425	***	0.14		0.153		0.011		—			
	Spearman's rho	-0.307	**	0.115		0.139		-0.1		—			
INTGa	*Pearson's r*	0.406	***	0.062		-0.083		0.034		-0.487	***	—	
	Spearman's rho	0.022		0.239	*	-0.076		0.542	***	-0.34	***	—	
ROA_win	*Pearson's r*	0.851	***	0.129		-0.498	***	-0.417	***	-0.535	***	0.274	**
	Spearman's rho	0.873	***	0.174		-0.489	***	-0.512	***	-0.451	***	0.01	

Note. * p < .05, ** p < .01, *** p < .001

Table L4: SME&S correlation matrix

Variable	*test*	GROSS	CAT	R&Dg	SG&A*	PPEa	INTGa
GROSS	*Pearson's r*	—					
	Spearman's rho	—					
CAT	*Pearson's r*	-0.420 ***	—				
	Spearman's rho	-0.353 ***	—				
R&Dg	*Pearson's r*	-0.062	-0.227 ***	—			
	Spearman's rho	0.004	-0.205 ***	—			
SG&A*	*Pearson's r*	-0.262 ***	0.072	0.486 ***	—		
	Spearman's rho	-0.241 ***	-0.005	0.316 ***	—		
PPEa	*Pearson's r*	-0.518 ***	0.162 **	-0.284 ***	-0.11	—	
	Spearman's rho	-0.436 ***	0.209 ***	-0.334 ***	-0.094	—	
INTGa	*Pearson's r*	-0.019	0.385 ***	0.109	0.074	-0.419 ***	—
	Spearman's rho	0.047	0.311 ***	0.241 ***	0.037	-0.416 ***	—
ROA_win	*Pearson's r*	0.477 ***	0.058	-0.619 ***	-0.717 ***	-0.069	-0.106
	Spearman's rho	0.512 ***	0.129 *	-0.547 ***	-0.737 ***	-0.061	-0.037

Note. * p < .05, ** p < .01, *** p < .001

Table L5: Model-IDM regression results

Model-IDM Fit Measures

Sector	R^2	Adj. R^2	Overall Model Test: F	df1	df2	p
IDM	0.727	0.714	55.6	9	188	< .001

Model-IDM Coefficients - ROA_win

Predictor	Estimate	SE	t	p	Stand. Estimate	Stand. Estimate 95% Conf. Interval: Lower	Upper
Independent variables							
Intercept [a]	-0.039	0.027	-1.450	0.149			
GROSS	0.389	0.037	10.567	< .001	0.564	0.459	0.669
CAT	0.025	0.008	3.012	0.003	0.150	0.052	0.248
R&Dg	-0.057	0.019	-3.012	0.003	-0.184	-0.304	-0.063
SG&A*	-0.132	0.026	-5.017	< .001	-0.340	-0.474	-0.206
PPEa	0.033	0.041	0.797	0.427	0.059	-0.087	0.204
INTGa	-0.127	0.027	-4.663	< .001	-0.343	-0.488	-0.198
Dummy variables							
REGION:							
East Asia – USA	0.001	0.009	0.118	0.906	0.012	-0.188	0.212
Europe – USA	0.012	0.011	1.122	0.263	0.144	-0.109	0.396
SIZE:							
Mid-sized firm – Big firm	-0.043	0.010	-4.246	< .001	-0.506	-0.741	-0.271

[a] Represents grand mean[27]

[27] Dummy coding and Simple coding are two ways to handle categories in statistical models.
Dummy coding (Reference level) involves leaving out one category as the reference and comparing all other categories against this reference. The results show how different each category is from the reference, based on the overall average of the data.
Simple coding (Grand mean), on the other hand, also picks a reference category, but it only compares each of the other categories directly to this reference. The results show the difference between each category and the reference specifically, without considering the overall average.

Table L6: Model-Fabless regression results

Model-Fabless Fit Measures							
				Overall Model Test			
Sector	**R^2**	**Adj. R^2**		**F**	**df1**	**df2**	**p**
Fabless	0.650	0.638		56.1	9	272	< .001

Model-Fabless Coefficients - ROA_win					**Stand. Estimate 95% Conf. Interval**		
Predictor	**Estimate**	**SE**	**t**	**p**	**Stand. Estimate**	**Lower**	**Upper**
Independent variables							
Intercept [a]	0.029	0.024	1.186	0.237			
GROSS	0.210	0.026	8.075	< .001	0.368	0.278	0.457
CAT	0.066	0.011	6.214	< .001	0.320	0.218	0.421
R&Dg	-0.130	0.018	-7.343	< .001	-0.413	-0.524	-0.303
SG&A*	-0.024	0.022	-1.105	0.270	-0.070	-0.193	0.054
PPEa	0.047	0.072	0.657	0.512	0.031	-0.061	0.122
INTGa	-0.181	0.025	-7.134	< .001	-0.372	-0.475	-0.270
Dummy variables							
REGION:							
East Asia – USA	0.046	0.010	4.419	< .001	0.466	0.258	0.673
Europe – USA	0.047	0.013	3.519	< .001	0.475	0.209	0.741
SIZE:							
Mid-sized firm – Big firm	-0.035	0.009	-3.891	< .001	-0.355	-0.534	-0.175

[a] Represents grand mean

Table L7: Model-Foundry regression results

Model-Foundry Fit Measures							
				Overall Model Test			
Sector	**R^2**	**Adj. R^2**		**F**	**df1**	**df2**	**p**
Foundry	0.899	0.899		83.3	9	84	< .001

Model-Foundry Coefficients - ROA_win					**Stand. Estimate 95% Conf. Interval**		
Predictor	**Estimate**	**SE**	**t**	**p**	**Stand. Estimate**	**Lower**	**Upper**
Independent variables							
Intercept [a]	-0.066	0.020	-3.324	0.001			
GROSS	0.480	0.029	16.541	< .001	0.818	0.720	0.916
CAT	0.085	0.011	7.647	< .001	0.346	0.256	0.435
R&Dg	-0.010	0.018	-0.559	0.577	-0.036	-0.162	0.091
SG&A*	-0.008	0.020	-0.414	0.680	-0.027	-0.159	0.104
PPEa	-0.194	0.023	-8.407	< .001	-0.379	-0.469	-0.290
INTGa	-0.215	0.059	-3.635	< .001	-0.189	-0.292	-0.086
Dummy variables							
REGION:							
East Asia – USA	0.026	0.012	2.214	0.030	0.309	0.032	0.586
Europe – USA	0.023	0.017	1.308	0.194	0.271	-0.141	0.682
SIZE:							
Mid-sized firm – Big firm	-0.005	0.009	-0.587	0.559	-0.064	-0.279	0.152

[a] Represents grand mean

Table L8: Model-SME&S regression results

Model-SME&S Fit Measures					**Overall Model Test**		
Sector	**R^2**	**Adj. R^2**		**F**	**df1**	**df2**	**p**
SME&S	0.775	0.768		107	9	279	< .001

Model-SME&S Coefficients - ROA_win						**Stand. Estimate 95% Conf. Interval**	
Predictor	**Estimate**	**SE**	**t**	**p**	**Stand. Estimate**	**Lower**	**Upper**
Independent variables							
Intercept [a]	0.083	0.022	3.732	< .001			
GROSS	0.236	0.030	7.790	< .001	0.327	0.245	0.410
CAT	0.052	0.008	6.593	< .001	0.254	0.178	0.330
R&Dg	-0.141	0.017	-8.453	< .001	-0.322	-0.397	-0.247
SG&A*	-0.162	0.014	-11.371	< .001	-0.457	-0.536	-0.378
PPEa	-0.101	0.024	-4.252	< .001	-0.184	-0.269	-0.099
INTGa	-0.122	0.023	-5.259	< .001	-0.199	-0.273	-0.124
Dummy variables							
REGION:							
East Asia – USA	-0.001	0.007	-0.160	0.873	-0.014	-0.183	0.155
Europe – USA	0.009	0.007	1.212	0.227	0.101	-0.063	0.264
SIZE:							
Mid-sized firm – Big firm	-0.015	0.006	-2.483	0.014	-0.177	-0.317	-0.037

[a] Represents grand mean

Table L9: Assumption check – Normality test

Shapiro-Wilk test	**IDM**	**Fabless**	**Foundry**	**SME&S**
Statistic	0.931	0.968	0.968	0.909
p	< .001	< .001	0.021	< .001

Table L10: Assumption check – Autocorrelation test

Durbin–Watson test	**IDM**	**Fabless**	**Foundry**	**SME&S**
Autocorrelation	-0.125	0.023	0.184	0.068
DW Statistic	2.25	1.95	1.59	1.84
p	0.076	0.640	0.056	0.188

Table L11: Assumption check – Collinearity Statistics

Collinearity Statistics	**IDM**		**Fabless**		**Foundry**		**SME&S**	
	VIF	**Tolerance**	**VIF**	**Tolerance**	**VIF**	**Tolerance**	**VIF**	**Tolerance**
GROSS	1.4	0.714	1.27	0.788	1.43	0.7	1.48	0.676
CAT	1.3	0.768	1.43	0.697	1.3	0.766	1.35	0.738
R&Dg	1.6	0.625	1.57	0.637	1.83	0.545	1.34	0.747
SG&A*	1.78	0.562	1.75	0.571	1.91	0.522	1.41	0.708
PPEa	1.93	0.518	1.29	0.774	1.3	0.767	1.52	0.657
INTGa	1.93	0.518	1.45	0.687	1.5	0.667	1.33	0.752
REGN	1.12	0.897	1.2	0.83	1.5	0.666	1.17	0.853
SIZE	1.21	0.828	1.21	0.829	1.35	0.74	1.2	0.83

Table L12: Dummy variables detailed interpretation

Dummy variables
REGION variable set-up
• Region 1 (A): This is indicated by (1, 0) — the first dummy variable is 1 and the second is 0. This refers to Europe compared to the USA (Europe-USA). • Region 2 (B): This is indicated by (0, 1) — the first dummy variable is 0 and the second is 1. This refers to East Asia compared to the USA (East Asia-USA). • Region 3: This is the baseline category indicated by (0, 0), where both dummy variables are 0. It refers to the USA since both the Europe and East Asia dummy variables are compared to the USA. The dummy variables for REGION, namely (East-Asia-USA) and (Europe-USA) surprisingly are mostly insignificant at 0.05 for most of the models. A possible reason is that strategic positioning and operational setups in these sectors could be sufficiently diversified across regions, balancing out regional advantages or disadvantages in the global market. Notable excepts are the positive influence of both dummies in Fabless and the positive dummy of (East-Asia – USA) for Foundries. A possible explanation is that Fabless companies in these regions may offer more specialized markets or favourable conditions such as better intellectual property protection, more advanced technological infrastructure, or supportive government policies, which could facilitate better performance for Fabless firms. The positive impact of the East-Asia-USA dummy on Foundries underlines the central role that East Asia plays in the global semiconductor manufacturing landscape. This region, especially Taiwan, South Korea, and to a lesser extent China, is home to some of the world's largest and most efficient semiconductor foundries.
FMSIZE variable set-up
• **Big Firms (Baseline):** This is the reference group, indicated by (0). • **Mid-sized Firm:** This group is indicated by (1). The dummy variable for firm size (FMSIZE) shows no significant impact at an $\alpha = 0.05$ in the Foundry sector. However, it has a strong negative effect in the IDM sector, a substantial negative effect in the Fabless sector, and a moderate negative impact in the SME&S sector. This could be because the foundry business is generally dominated by a few large players. The sector is structured in such a way that all active foundries need to operate at a certain scale to remain viable. The strong negative impact of being smaller than Big firms in the IDM sector suggests significant disadvantages for smaller players. IDMs benefit from economies of scale, integrated operations, and extensive R&D capabilities, which are more effectively realized in larger operations. Smaller IDMs might struggle with capital intensity. Similarly, in the Fabless segment, smaller sizes relative to Big firms also have a strong negative impact. This reflects challenges in competing with larger entities that can command more resources for R&D and securing strategic partnerships with Foundries. The medium negative impact for smaller SME&S indicates that while size does affect their performance, it might not be as critical as in IDMs or Fabless companies. SME&S can operate successfully in niche markets or specialized services where the scalability of operations isn't as crucial, but still, larger size could confer benefits like better access to capital, more significant market reach, and greater resilience to market fluctuations.

Appendix M: Python and R Scripts

Script M1: Extracting Company names and Headquarters (Van Rossum & Drake, 2009)

```
# 1.1 Extracting Company names and Headquarters:
                    # Check "Manual path full" that responds to "Path full"
                    # Check that "Manual path full" are only VALUES and not FORMULA
                    # Before running the code clean column E as it won't overwrite it

# Import ALL the libraries
  import openpyxl
  from collections import OrderedDict
  from openpyxl.utils import get_column_letter

main_file_path = r'C:\Users\...\Panel Data (100)_v5\Panel_Data_v5.xlsx'
main_sheet_name = 'Statements'

main_workbook = openpyxl.load_workbook(main_file_path)
main_sheet = main_workbook[main_sheet_name]

row_index = 2                                                    # Initialize row index
while True:
    file_path = main_sheet[f'D{row_index}'].value
    if not file_path:
        break
    try:
        if file_path.lower().endswith('.xlsx'):                  # Open the corresponding file
            file_workbook = openpyxl.load_workbook(file_path, read_only=True)
        else:
            file_workbook = openpyxl.open(file_path, read_only=True)
        file_sheet = file_workbook['Financial Summary']
        cell_b2_value = file_sheet['B2'].value           # Read the value from cell B2 (Company)…
        cell_b4_value = file_sheet['B4'].value           # …and B4 (Country of Headquarters)
        file_workbook.close()
        main_sheet[f'E{row_index}'].value = cell_b2_value # Write the value to the corresponding cell
        main_sheet[f'F{row_index}'].value = cell_b4_value
    except FileNotFoundError:
        print(f"File not found: {file_path}")
    except Exception as e:
        print(f"Error processing file {file_path}: {e}")
    row_index += 1                                    # Move to the next row
main_workbook.save(main_file_path)          # Save changes to the main workbook
main_workbook.close()
```

Script M2: Extract paths and register (Van Rossum & Drake, 2009)

```
# 1.2 Extract paths and register
"""
          In the excel sheet "paths" that will be create,
          apply the =A1&B1&C1&D1 formula where A1 = r', B1 = path, C1 = ', D1 = '
          in order to then just copy it in the path_file directly
          Note: at D1 = , don't include the comma at the last row
"""


main_file_workbook = openpyxl.load_workbook(main_file_path)
sheet_Statements = main_file_workbook['Statements']
file_paths_column = sheet_Statements['D'][1:]     # Get all values from column D starting from row 2


# Use OrderedDict to maintain the order while eliminating duplicates
unique_file_paths = list(OrderedDict.fromkeys(cell.value for cell in file_paths_column if cell.value))
paths_sheet = main_file_workbook.create_sheet(title='paths', index=0)
for index, path in enumerate(unique_file_paths, start=1): # Write unique file paths to sheet "paths"…
    paths_sheet.cell(row=index, column=2, value=path)     # …in column B starting from B1


main_file_workbook.save(main_file_path)
main_file_workbook.close()
```

Script M3: Extract Statement data and Period end date (Van Rossum & Drake, 2009)

```
# 1.3 Extract Statement data and Period end date
      # Note: copy-paste the paths from the "path" sheet in the file_path


def process_file(file_path):
    data_workbook = openpyxl.load_workbook(file_path)
    balance_sheet = data_workbook['Balance Sheet']
```

```python
    values_to_record_1 = []
    values_to_record_2 = []


    for i in range(2, 12):                                  # Assuming values to be read from column B (2) to K (11)
        value_1 = balance_sheet.cell(row=11, column=i).value
        values_to_record_1.append(value_1)
        value_2 = balance_sheet.cell(row=12, column=i).value
        values_to_record_2.append(value_2)

    data_workbook.close()
    return values_to_record_1, values_to_record_2


main_file_workbook = openpyxl.load_workbook(main_file_path)
sheet_Statements = main_file_workbook['Statements']
row_num = 2                                                 # Starting row in main Excel file, sheet "Statements"
file_paths = [
    r'C:\Users\...\Annual Financial Statements (sample)\Samsung Electronics Co Ltd (005930.KS).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Intel Corp (INTC.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Taiwan Semiconductor Manufacturing Co Ltd (2330.TW).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\SK Hynix Inc (000660.KS).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Broadcom Inc (AVGO.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Advanced Micro Devices Inc (AMD.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Micron Technology Inc (MU.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Analog Devices Inc (ADI.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Qualcomm Inc (QCOM.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Semiconductor Manufacturing International Corp
(0981.HK).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\ASML Holding NV (ASML.AS).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\NVIDIA Corp (NVDA.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Texas Instruments Inc (TXN.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Applied Materials Inc (AMAT.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Infineon Technologies AG (IFXGn.DE).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\NXP Semiconductors NV (NXPI.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Marvell Technology Inc (MRVL.OQ).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Renesas Electronics Corp (6723.T).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\STMicroelectronics NV (STMMI.MI).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\MediaTek Inc (2454.TW).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Lam Research Corp (LRCX.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\United Microelectronics Corp (2303.TW).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Tokyo Electron Ltd (8035.T).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Microchip Technology Inc (MCHP.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\KLA Corp (KLAC.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Coherent Corp (COHR.K).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\ON Semiconductor Corp (ON.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Entegris Inc (ENTG.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Synopsys Inc (SNPS.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Skyworks Solutions Inc (SWKS.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Rohm Co Ltd (6963.T).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Amkor Technology Inc (AMKR.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Qorvo Inc (QRVO.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Wolfspeed Inc (WOLF.K).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Sino-American Silicon Products Inc (5483.TWO).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Globalwafers Co Ltd (6488.TWO).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Advantest Corp (6857.T).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\AT & S Austria Technologie & Systemtechnik AG (ATSV.VI).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Siltronic AG (WAFGn.DE).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\ASM International NV (ASMI.AS).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Vishay Intertechnology Inc (VSH).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Realtek Semiconductor Corp (2379.TW).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Disco Corp (6146.T).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Vanguard International Semiconductor Corp (5347.TWO).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Novatek Microelectronics Corp (3034.TW).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\ASMPT Ltd (0522.HK).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Soitec SA (SOIT.PA).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Semtech Corp (SMTC.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\WIN Semiconductors Corp (3105.TWO).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Silicon Laboratories Inc (SLAB.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Cirrus Logic Inc (CRUS.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Monolithic Power Systems Inc (MPWR.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Ultra Clean Holdings Inc (UCTT.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Lasertec Corp (6920.T).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Onto Innovation Inc (ONTO.K).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Himax Technologies Inc (HIMX.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Kulicke and Soffa Industries Inc (KLIC.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\SMART Global Holdings Inc. (SGH.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\MACOM Technology Solutions Holdings Inc (MTSI.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\ChipMOS Technologies Inc (8150.TW).xlsx',
```

```
    r'C:\Users\...\Annual Financial Statements (sample)\Photronics Inc (PLAB.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\X Fab Silicon Foundries EV (XFAB.PA).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\ACM Research Inc (ACMR.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Cohu Inc (COHU.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\BE Semiconductor Industries NV (BESI.AS).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Alpha and Omega Semiconductor Ltd (AOSL.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Formosa Sumco Technology Corp (3532.TW).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Silergy Corp (6415.TW).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Veeco Instruments Inc (VECO.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Axcelis Technologies Inc (ACLS.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\FormFactor Inc (FORM.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Aixtron SE (AIXGn.DE).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Silicon Motion Technology Corp (SIMO.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Power Integrations Inc (POWI.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\MARUWA Co Ltd (5344.T).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Lattice Semiconductor Corp (LSCC.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Nordic Semiconductor ASA (NOD.OL).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\U Blox Holding AG (UBXN.S).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Ambarella Inc (AMBA.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Global Unichip Corp (3443.TW).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Alchip Technologies Ltd (3661.TW).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Melexis NV (MLXS.BR).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\AXT Inc (AXTI.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\nLIGHT Inc (LASR.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\CEVA Inc (CEVA.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\PDF Solutions Inc (PDFS.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\eMemory Technology Inc (3529.TWO).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\inTest Corp (INTT.K).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Pixelworks Inc (PXLW.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Aehr Test Systems (AEHR.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Everspin Technologies Inc (MRAM.O).xlsx',
    r'C:\Users\...\Annual Financial Statements (sample)\Quicklogic Corp (QUIK.O).xlsx'
   ]
""" END file_path"""
for file_path in file_paths:      # Process each file and record values in main Excel file, sheet "Statements"
    values_to_record_1, values_to_record_2 = process_file(file_path)
    for i, value in enumerate(values_to_record_1, start=row_num):
        sheet_Statements.cell(row=i, column=8).value = value
    for i, value in enumerate(values_to_record_2, start=row_num):
        sheet_Statements.cell(row=i, column=9).value = value
    # Move to the next row
    row_num += max(len(values_to_record_1), len(values_to_record_2))
main_file_workbook.save(main_file_path)
main_file_workbook.close()
```

Script M4: Extract variables (Income statement) (Van Rossum & Drake, 2009)

```
# 2.1 Extract variables (PnL)
"""
    The code assumes that I won't have more than 30 accounts (columns E to AA).
    To add more, just adjust the range accordingly.
"""
def find_row_by_variable(sheet, variable_name):
    for row_num in range(1, sheet.max_row + 1):
        if sheet.cell(row=row_num, column=1).value == variable_name:
            return row_num
    return None


def process_file(file_path, sheet_Variables, start_row):
    data_workbook = openpyxl.load_workbook(file_path)
    data_sheet = data_workbook['Income Statement']
    for col_offset in range(0, 30):                          # Assuming no more than 30 accounts
        variable_name = sheet_Variables.cell(row=1, column=7 + col_offset).value
        variable_row = find_row_by_variable(data_sheet, variable_name)
        if variable_row is not None:
            for row_num in range(start_row, start_row + 10):
                year_to_search = sheet_Variables.cell(row=row_num, column=4).value
                if year_to_search is not None:
                    for col_num in range(2, 12):
                        header_year = data_sheet.cell(row=11, column=col_num).value
                        if header_year == year_to_search:
                            # Record the value in the Variables sheet
                            value_to_record = data_sheet.cell(row=variable_row, column=col_num).value
                            sheet_Variables.cell(row=row_num, column=7 + col_offset).value = value_to_record
    data_workbook.close()


main_file_workbook = openpyxl.load_workbook(main_file_path)
```

```
sheet_Variables = main_file_workbook['Income Statements']

row_num = 2                         # Starting row in main Excel file, sheet "Statements"
start_row = 2                       # Initialize the starting row for data recording
for file_path in file_paths:        # Process each data file and record values in the main workbook
    process_file(file_path, sheet_Variables, start_row)
    start_row += 10                 # Move to the next Company/set of 10 rows
main_file_workbook.save(main_file_path)
main_file_workbook.close()
```

Script M5: Extract variables (Balance sheet)

```
# Extract variables (Balance sheet)
"""
The code assumes that I won't have more than 30 variables (columns E to AA).
To add more, just adjust the range accordingly.

Handle skipping the specified variables ("Total Assets" and "Total Liabilities")…
…as they repeat twice and the 1st time is the title of a section
"""
def find_row_by_variable(sheet, variable_name):
    variables_to_skip = ["Total Assets", "Total Liabilities"] # List of variables to skip on the first occurrence
    # Dictionary to track skipping status for each variable
    skip_first_occurrence = {variable: False for variable in variables_to_skip}
    for row_num in range(1, sheet.max_row + 1):
        cell_value = sheet.cell(row=row_num, column=1).value
        if cell_value == variable_name:                          # Check if the current variable needs to be
skipped
            if variable_name in variables_to_skip and not skip_first_occurrence[variable_name]:
                skip_first_occurrence[variable_name] = True      # Skip the first occurrence
            elif variable_name in variables_to_skip and skip_first_occurrence[variable_name]:
                return row_num                                   # Return the row number for the second occurrence
            elif variable_name not in variables_to_skip:
                return row_num                                   # Return the row number for variables
    return None                                                    … not in the skip list

def process_file(file_path, sheet_Variables, start_row):
    data_workbook = openpyxl.load_workbook(file_path)
    data_sheet = data_workbook['Balance Sheet']
    for col_offset in range(0, 100):  # Assuming no more than 30 accounts
        variable_name = sheet_Variables.cell(row=1, column=7 + col_offset).value
        variable_row = find_row_by_variable(data_sheet, variable_name)
        if variable_row is not None:
            for row_num in range(start_row, start_row + 10):
                year_to_search = sheet_Variables.cell(row=row_num, column=4).value
                if year_to_search is not None:
                    for col_num in range(2, 12):
                        header_year = data_sheet.cell(row=11, column=col_num).value
                        if header_year == year_to_search:
                            # Record the value in the Variables sheet
                            value_to_record = data_sheet.cell(row=variable_row, column=col_num).value
                            sheet_Variables.cell(row=row_num, column=7 + col_offset).value = value_to_record
    data_workbook.close()
sheet_Variables = main_file_workbook['Balance Sheets']
row_num = 2               # Starting row in main Excel file, sheet "Statements"
file_paths = file_paths
start_row = 2             # Initialize the starting row for data recording

for file_path in file_paths: # Process each data file and record values in the main workbook
    process_file(file_path, sheet_Variables, start_row)
    start_row += 10          # Move to the next set of 10 rows
main_file_workbook.save(main_file_path)
main_file_workbook.close()
```

Script M6: Detection of companies with more than two years of missing statements (Van Rossum & Drake, 2009)

```
import pandas as pd
file_path = r'C:\Users\...\Data\Exclussion\Panel_Data_e0.xlsx'
sheets = ['Income Statements', 'Balance Sheets', 'Financial Summaries']

def process_sheet(sheet_name):                                  # Function to process each sheet
    df = pd.read_excel(file_path, sheet_name=sheet_name)        # Read the sheet into a DataFrame
    num_chunks = len(df) // 10                        # Get the number of chunks (assuming each chunk is 10 rows)
    for i in range(num_chunks):                                 # Iterate over the chunks
```

```
        start_row = i * 10
        end_row = start_row + 10                        # Calculate the start and end row for the current chunk
        chunk = df.iloc[start_row:end_row]              # Extract the chunk
        missing_a = chunk['Company'].isnull().any()     # Check if there is any missing data in column 'A' or 'F'
        missing_f_count = chunk['Fiscal year'].isnull().sum()


        # Print messages and drop rows if needed
        if not missing_a and missing_f_count == 0:      # All data is present, do nothing
            continue
        elif not missing_a and 0 < missing_f_count < 3:      # Partial data missing
            print(f"Company {chunk['Company'].iloc[0]} has {missing_f_count} years missing.")
        elif not missing_a and missing_f_count >= 3:        # A lot of data missing, print message and drop rows
            print(f"Company {chunk['Company'].iloc[0]} has {missing_f_count} years missing. TERMUNATION!")
            df.drop(df.index[start_row:end_row], inplace=True)

    # Save the modified DataFrame back to the Excel file
    with pd.ExcelWriter(file_path, engine='openpyxl', mode='a', if_sheet_exists='replace') as writer:
        df.to_excel(writer, sheet_name=sheet_name, index=False)

for sheet in sheets:                                     # Process each sheet
    process_sheet(sheet)
```

Script M7: Exclude specified companies (Van Rossum & Drake, 2009)

```
# Insert the names of the companies that were specified for exclusion by previous code:
exclude_companies = ['MagnaChip Semiconductor Corp (MX)',
                     'Atomera Inc (ATOM.O)',
                     'Sequans Communications SA (SQNS.K)']       # List of companies to exclude


def process_sheet(sheet_name, file_path):                         # Function to process each sheet and print deleted
rows
    df = pd.read_excel(file_path, sheet_name=sheet_name)
    rows_to_delete = df[df.iloc[:, 0].isin(exclude_companies)]    # Assuming the first column contains the company names
                                                                  # Find rows to delete based on the first column
    if not rows_to_delete.empty:                                  # If there are rows to delete, print them
        print(f"Deleted rows from '{sheet_name}':")
        print(rows_to_delete)
        print("\n")
    df = df[~df.iloc[:, 0].isin(exclude_companies)]               # Del rows if the 1st col match any exclude_companies
    with pd.ExcelWriter(file_path, engine='openpyxl', mode='a', if_sheet_exists='replace') as writer:
        df.to_excel(writer, sheet_name=sheet_name, index=False)
for sheet in sheets:
    process_sheet(sheet, file_path)
```

Script M8: Winsorization of variables (R Core Team, 2021).

```
# Outlier analysis by 1 Variable (Winsorizing and Adjustments)
"
   Name the file _data and then proceed
   Don't forget to rename the file at the end of the process to the respective variable name
"
    rm(list = ls())                                                # Remove all the existing objects & load libraries
    library(tidyverse)
    library(readxl)
    library(openxlsx)
    library(ggplot2)
    library(palmerpenguins)
    library(DescTools)
    library(dplyr)
    library(scales)
    library(RColorBrewer)

setwd("C:/Users/.../Thesis/Data/Outlier Analysis/Winsorizing FINAL") # Setting the working directory
getwd()
data <- read_excel("_data.xlsx", sheet = "Windsorizing")
names(data)                                                        # Check column names
#----------------------------------------------------------------------------------------------------------
sum(is.na(data))                   # Find Missing data if any:
start_column <- 6                  # Use the column index directly, alternatively, use the name of the column:
                                   # start_column <- which(names(data) == "Revenue from Business Activities")
data[, start_column:ncol(data)][is.na(data[, start_column:ncol(data)])] <- 0  #  Fill the missing values with 0
sum(is.na(data))                   # Verify the operation by checking for missing values
# Write the modified 'data' back to a new Excel file or overwrite the existing one
write.xlsx(data, file = "_data.xlsx", sheetName = "Windsorizing", overwrite = TRUE)
print("File successfully modified")
#----------------------------------------------------------------------------------------------------------
```

```r
#____________________________________Outlier Analysis_____________________________________________________
#   Using BoxPlot to detect the presence of outliers
data <- read_excel("_data.xlsx", sheet = "Windsorizing")
   data <- data %>%                                                   # Add a temporary column for plotting
       mutate(Temp_Column = .[[6]])
ggplot(data, aes(x = factor(`Fiscal year`), y = Temp_Column)) +   # Plot using this temporary column
    geom_boxplot(width = 0.2) +
    theme_classic()
data <- select(data, -Temp_Column)                                # Remove the temporary column after plotting
#-----------------------------------------------------------------------------------------------------------


#   Create Winsorized variable → Winsorizing the variable at the 5th and 95th percentiles by groups
win_data <- data %>%
   group_by(`Fiscal year`) %>%
    mutate(across(.cols = 5, .fns = ~DescTools::Winsorize(.x, probs = c(0.05, 0.95)), .names = "Winsorized_variable"))
%>%
       ungroup()
ggplot(win_data, aes(x = factor(`Fiscal year`), y = Winsorized_variable)) +    # Winsorized Data Plot
    geom_boxplot(fill = "#92C5DE") +
     labs(title = "Winsorized Variable",
         x = "Fiscal Year",
         y = "Ratios") +
     theme_minimal() +
         theme(axis.text.x = element_text(angle = 45, hjust = 1))

#   Modify Excel: Add the Winsorized_variable columns from 'win_data' to the 'data' dataframe
     data$Winsorized_variable <- win_data$Winsorized_variable
#   Modify Excel: Add a column to indicate if an Winsorization was made
     data$Winsorize_Made <- ifelse(data[[6]] != data$Winsorized_variable, "Yes", "No")
#   Write the updated 'data' dataframe back to an Excel file
filePath <- "_data.xlsx"                                         # Load or create the Excel workbook
if (file.exists(filePath)) {
    wb <- loadWorkbook(filePath)
} else {
    wb <- createWorkbook()
    addWorksheet(wb, "Windsorizing")
}
    sheetName <- "Windsorizing"
columnNames <- c("Winsorized_variable", "Winsorize_Made")        # Column names and their respective start columns
startCols <- c(7, 8)
#   Write column names in row 1
writeData(wb, sheet = sheetName, x = matrix(columnNames, nrow = 1), startCol = startCols, startRow = 1, colNames = FALSE)
#   Write the data to the Excel file, placing each column in its designated column
writeData(wb, sheet = sheetName, x = data$Winsorized_variable, startCol = 7, startRow = 2, colNames = FALSE)
writeData(wb, sheet = sheetName, x = data$Winsorize_Made, startCol = 8, startRow = 2, colNames = FALSE)
saveWorkbook(wb, filePath, overwrite = TRUE)                      # Save the workbook
#-----------------------------------------------------------------------------------------------------------


#  Adjust the winsorized dataset by replacing the remaining outliers with the highest/lowest non-outlier value
#  (i.e., the value at the upper quartile plus 1.5 times the interquartile range)
#  Adjusted function to replace outliers within a vector including both upper and lower bounds
replace_outliers_vector <- function(variable_vector) {            # Adjusted function to replace outliers within a…
    Q1 <- quantile(variable_vector, 0.25)                         # …vector including both upper and lower bounds
    Q3 <- quantile(variable_vector, 0.75)
    IQR <- Q3 - Q1
    upper_bound <- Q3 + 1.5 * IQR                                 # Define upper and lower bounds
    lower_bound <- Q1 - 1.5 * IQR
       return(ifelse(variable_vector > upper_bound, upper_bound,  # Replace outliers and return the modified vector
              ifelse(variable_vector < lower_bound, lower_bound, variable_vector)))
}
adjusted_data <- win_data %>%                                      # Apply the adjusted function to each fiscal year
   group_by(`Fiscal year`) %>%
   mutate(`Winsorized_variable` = replace_outliers_vector(`Winsorized_variable`)) %>%
   ungroup()                        # The `adjusted_data` now contains variables where outliers have been adjusted.
#-----------------------------------------------------------------------------------------------------------
# Visualize the results of outlier adjustment → Adjusted Winsorized Data Plot
ggplot(adjusted_data, aes(x = factor(`Fiscal year`), y = Winsorized_variable)) +
     geom_boxplot(fill = "#92C5DE") +
     labs(title = "Adjusted Winsorized Variable",
         x = "Fiscal Year",
         y = "Ratios") +
     theme_minimal() +
     theme(axis.text.x = element_text(angle = 45, hjust = 1))
#-----------------------------------------------------------------------------------------------------------


# Save changes: 'Winsorized_variable' in 'adjusted_data' is the column to rename to 'Adjusted_Winsorized_variable'
        adjusted_data$Adjusted_Winsorized_variable <- adjusted_data$Winsorized_variable
```

```r
# Save changes: Now, copying 'Winsorized_variable' from 'win_data' to 'adjusted_data', under the name
'Winsorized_variable'
        adjusted_data$Winsorized_variable <- win_data$Winsorized_variable
# Next, assess if an adjustment has been made.
# The comparison is between 'Adjusted_Winsorized_variable' (the newly adjusted values) and the original…
#  …'Winsorized_variable' values just copied from 'win_data'
adjusted_data$Adjustment_Made <- ifelse(adjusted_data$Adjusted_Winsorized_variable < adjusted_data$Winsorized_variable,
"Yes", "No")
# Finally, save the adjusted data to an Excel file
filePath <- "_data.xlsx"                        # Load the existing workbook or create a new one if it doesn't exist
if (file.exists(filePath)) {
    wb <- loadWorkbook(filePath)
} else {
    wb <- createWorkbook()
    addWorksheet(wb, "Windsorizing")
}
     sheetName <- "Windsorizing"
 columnNames <- c("Adjusted_Winsorized_variable", "Adjustment_Made") # Column names and their respective start columns
 startCols <- c(9, 10)
# Write column names in row 1
writeData(wb, sheet = sheetName, x = matrix(columnNames, nrow = 1), startCol = startCols, startRow = 1, colNames = FALSE)
# Writing 'Adjusted_Winsorized_variable' to column K and 'Adjustment_Made' to column L
writeData(wb, sheet = sheetName, x = adjusted_data$Adjusted_Winsorized_variable, startCol = 9, startRow = 2, colNames =
FALSE)
writeData(wb, sheet = sheetName, x = adjusted_data$Adjustment_Made, startCol = 10, startRow = 2, colNames = FALSE)
saveWorkbook(wb, filePath, overwrite = TRUE)                     # Save the workbook
#-------------------------------------------------------------------------------------------------------------
print("Note: Rename the file _data to respective variable name before proceeding to the next one")
print("__________END__________")
```

Script M9: χ2 test of independence/association (R Core Team, 2021).

```python
import pandas as pd
from scipy.stats import chi2_contingency


file_path = r'C:\Users\...\Thesis\Data\Chi-Square Test\Chi^2.xlsx'
sheet_name = "10.3"
data = pd.read_excel(file_path, sheet_name=sheet_name)
missing_values = data[['SIZE', 'YEAR', 'SECTR', 'REGN']].isnull().sum()
missing_values                                                            # Check for missing values


SIZE_YEAR_crosstab = pd.crosstab(data['SIZE'], data['YEAR'])              # Prepare the data for Chi^2 tests
SIZE_type_crosstab = pd.crosstab(data['SIZE'], data['SECTR'])
SIZE_REGN_crosstab = pd.crosstab(data['SIZE'], data['REGN'])
YEAR_type_crosstab = pd.crosstab(data['YEAR'], data['SECTR'])
YEAR_REGN_crosstab = pd.crosstab(data['YEAR'], data['REGN'])
type_REGN_crosstab = pd.crosstab(data['SECTR'], data['REGN'])


chi2_results = {                                                          # Perform Chi-squared tests
    'SIZE vs YEAR': chi2_contingency(SIZE_YEAR_crosstab),
    'SIZE vs SECTR': chi2_contingency(SIZE_type_crosstab),
    'SIZE vs REGN': chi2_contingency(SIZE_REGN_crosstab),
    'YEAR vs SECTR': chi2_contingency(YEAR_type_crosstab),
    'YEAR vs REGN': chi2_contingency(YEAR_REGN_crosstab),
    'SECTR vs REGN': chi2_contingency(type_REGN_crosstab)
}
chi2_summary = {test: (result[0], result[1]) for test, result in chi2_results.items()}
chi2_summary
```

Script M10: ANOVA - Analysis of Variance (Van Rossum & Drake, 2009)

```python
import pandas as pd
import statsmodels.api as sm
from statsmodels.formula.api import ols
from scipy.stats import levene, shapiro
import openpyxl


# Load the Excel file
file_path = 'C:/Users/.../ANOVA (Analysis of Variance)/i3_Imp.ratios.xlsx'
sheet_name = 'Imp.ratios (10.3)'


data = pd.read_excel(file_path, sheet_name=sheet_name)


# Replace spaces with double underscores in column names
```

```
data.columns = data.columns.str.replace(' ', '__')


# List of continuous variables starting from column E
continuous_vars = data.columns[4:]


# List of categorical variables
categorical_vars = ['Region', 'Sector', 'Size']


# Create a writer object to save results
output_file = 'C:/Users/.../ANOVA (Analysis of Variance)/i3_Imp.ratios_ANOVA_results.xlsx'
writer = pd.ExcelWriter(output_file, engine='openpyxl')


# Function to perform ANOVA and tests
def perform_anova(data, dependent_var, categorical_vars):
    formula = f'Q("{dependent_var}") ~ ' + ' + '.join([f'Q("{var}")' for var in categorical_vars]
                  + [f'Q("{x1}")*Q("{x2}")' for i, x1 in enumerate(categorical_vars) for x2 in categorical_vars[i+1:]])
    model = ols(formula, data=data).fit()
    anova_table = sm.stats.anova_lm(model, typ=2)

    # Levene's test for homogeneity of variances
    levene_test = levene(*[group[dependent_var].values for name, group in data.groupby(categorical_vars)])

    # Shapiro-Wilk test for normality
    shapiro_test = shapiro(model.resid)

    return anova_table, levene_test, shapiro_test


# Iterate over each continuous variable
for dependent_var in continuous_vars:
    anova_table, levene_test, shapiro_test = perform_anova(data, dependent_var, categorical_vars)

    # Sanitize the sheet name
    sanitized_sheet_name = dependent_var.replace('/', '_per_')
                                        .replace('\\', '_')
                                        .replace('?', '')
                                        .replace('*', '')
                                        .replace(':', '-')
                                        .replace('[', '')
                                        .replace(']', '')

    # Save the results to a new sheet
    anova_table.to_excel(writer, sheet_name=sanitized_sheet_name, startrow=0)

    levene_results = pd.DataFrame({'Test': ['Levene\'s Test'], 'F': [levene_test[0]], 'p-value': [levene_test[1]]})
    levene_results.to_excel(writer, sheet_name=sanitized_sheet_name, startrow=anova_table.shape[0] + 3, index=False)

    shapiro_results = pd.DataFrame({'Test': ['Shapiro-Wilk Test'], 'Statistic': [shapiro_test[0]], 'p-value':
[shapiro_test[1]]})
    shapiro_results.to_excel(writer, sheet_name=sanitized_sheet_name, startrow=anova_table.shape[0] + 7, index=False)


# Save and close the writer
writer.close()
```

Script M11: Kruskal-Wallis, Levene's test, and Shapiro-Wilk test - Tests for Assumptions and Distribution (Van Rossum & Drake, 2009)

```
mport pandas as pd
from scipy.stats import kruskal, levene, shapiro
import openpyxl


# Load the Excel file
file_path = 'C:/Users/.../ANOVA (Analysis of Variance)/i3_Imp.ratios.xlsx'
sheet_name = 'Imp.ratios (10.3)'


# Load the data
data = pd.read_excel(file_path, sheet_name=sheet_name)


# Clean up column names
data.columns = data.columns.str.replace(' ', '__').str.replace('.', '').str.replace('(', '').str.replace(')', '')


# List of continuous variables starting from column E
```

```
continuous_vars = data.columns[4:]


# List of categorical variables
categorical_vars = ['Region', 'Sector', 'Size']


# Create a writer object to save results
output_file = 'C:/Users/.../ANOVA (Analysis of Variance)/i3_Imp.ratios_Kruskal-Wallis_results.xlsx'
writer = pd.ExcelWriter(output_file, engine='openpyxl')


# Function to perform Kruskal-Wallis and other tests
def perform_kruskal_wallis(data, dependent_var, categorical_vars):
    # Kruskal-Wallis test for multiple groups
    groups = [group[dependent_var].values for name, group in data.groupby(categorical_vars)]
    kruskal_test = kruskal(*groups)

    # Levene's test for homogeneity of variances
    levene_test = levene(*groups)

    # Shapiro-Wilk test for normality
    shapiro_test = shapiro(data[dependent_var])

    return kruskal_test, levene_test, shapiro_test


# Iterate over each continuous variable
for dependent_var in continuous_vars:
    kruskal_test, levene_test, shapiro_test = perform_kruskal_wallis(data, dependent_var, categorical_vars)

    # Sanitize the sheet name
    sanitized_sheet_name = dependent_var.replace('/', '_per_')
                                        .replace('\\', '_')
                                        .replace('?', '')
                                        .replace('*', '')
                                        .replace(':', '-')
                                        .replace('[', '')
                                        .replace(']', '')

    # Save the results to a new sheet
    kruskal_results = pd.DataFrame({'Test': ['Kruskal-Wallis Test'], 'Statistic': [kruskal_test[0]], 'p-value':
[kruskal_test[1]]})
    kruskal_results.to_excel(writer, sheet_name=sanitized_sheet_name, startrow=0, index=False)

    levene_results = pd.DataFrame({'Test': ['Levene\'s Test'], 'Statistic': [levene_test[0]], 'p-value': [levene_test[1]]})
    levene_results.to_excel(writer, sheet_name=sanitized_sheet_name, startrow=4, index=False)

    shapiro_results = pd.DataFrame({'Test': ['Shapiro-Wilk Test'], 'Statistic': [shapiro_test[0]], 'p-value':
[shapiro_test[1]]})
    shapiro_results.to_excel(writer, sheet_name=sanitized_sheet_name, startrow=8, index=False)


# Save and close the writer
writer.close()
```

Bibliography

Abdulhafedh, A. (2022). Modeling Vehicle Crash Frequency When Multicollinearity Exists in Vehicle Crash Data: Ridge Regression versus Ordinary Least Squares Linear Regression. *Open Access Library Journal*, *9*(6), 1–17. https://doi.org/10.4236/OALIB.1108873

Accenture. (2021). *Top 5 Trends to Impact Semiconductor Industry*. https://www.accenture.com/us-en/blogs/high-tech/top-5-trends-to-impact-the-semi-industry-in-2022

Adzan, M., Singh, A., & Mannesson, H. (2017). *Cities grow smarter through innovative semiconductor technologies*. https://www.ti.com/lit/wp/sszy031/sszy031.pdf?ts=1712852350254&ref_url=https%253A%252F%252Fwww.google.com%252F

Applied Energy Systems, Inc. (2023, October). *Navigating Complexities: The Semiconductor Shortage's Effect on the Auto Industry*. https://www.appliedenergysystems.com/semiconductor-shortages-effect-on-auto-industry/

Banerjee, A., Chitnis, U. B., Jadhav, S. L., Bhawalkar, J. S., & Chaudhury, S. (2009). Hypothesis testing, type I and type II errors. *Industrial Psychiatry Journal*, *18*(2), 127. https://doi.org/10.4103/0972-6748.62274

Berk, J., & DeMarzo, P. (2020). *Corporate Finance*.

Burja, C. (2011a). Factors Influencing The Companies' Profitability. *Annales Universitatis Apulensis Series Oeconomica*, *2*(13), 215–224. https://doi.org/10.29302/OECONOMICA.2011.13.2.3

Burja, C. (2011b). Factors Influencing The Companies' Profitability. *Annales Universitatis Apulensis Series Oeconomica*, *2*(13), 215–224. https://doi.org/10.29302/OECONOMICA.2011.13.2.3

Chang, B. G., & Wu, K. S. (2022). Concave Effect of Financial Flexibility on Semiconductor Industry Performance: Quantile Regression Approach. *Technological and Economic Development of Economy*, *28*(4), 948–978. https://doi.org/10.3846/TEDE.2022.16622

Companies Marketcap. (2024, February). *Largest semiconductor companies by market cap*. https://companiesmarketcap.com/semiconductors/largest-semiconductor-companies-by-market-cap/

Consumer Electronics Association (CEA). (2014). *Consumer Electronics Industry Revenues to Reach All-Time High in 2014, Projects CEA's Semi-Annual Sales and Forecasts Report | Business Wire*. https://www.businesswire.com/news/home/20140715006181/en/Consumer-Electronics-Industry-Revenues-to-Reach-All-Time-High-in-2014-Projects-CEA%E2%80%99s-Semi-Annual-Sales-and-Forecasts-Report

Counterpoint. (2024, January). *Global Semiconductor Revenues Decline 8.8% in 2023; AI Promises Gains in 2024*. https://www.counterpointresearch.com/insights/global-semiconductor-revenues-decline-8-8-in-2023-ai-promises-gains-in-2024/

Deloitte. (2024). 2024 global semiconductor industry outlook. *Deloitte*. https://www2.deloitte.com/content/dam/Deloitte/us/Documents/technology-media-telecommunications/us-tmt-semiconductor-industry-outlook-2024.pdf

Doshi, B. (2010). *Handling Missing Values in Data Mining*. https://www.semanticscholar.org/paper/Handling-Missing-Values-in-Data-Mining-Doshi/3817b208fe1f40891cc661ea0db80c8fccc56b70

Dumbravă, M. (2010). Model for the Analysis of the Company Performance. *Theoretical and Applied Economics*, *XVII*(8), 103–116. https://www.researchgate.net/publication/46567550_Model_for_the_Analysis_of_the_Company_Performance

European Commision. (2023). *European Chips Act - European Commission*. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-chips-act_en

European Commission. (2022). *European Chips Act*. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-chips-act_en

Fernàndez, J. R., Etzel, C., Beasley, T. M., Shete, S., Amos, C. I., & Allison, D. B. (2002). Improving the power of sib pair quantitative trait loci detection by phenotype Winsorization. *Human Heredity*, *53*(2), 59–67. https://doi.org/10.1159/000057984

Filatov, E. A. (2018). *Integral Factor Analysis Of Financial Profitability By Filatov's Method*. 356–363. https://doi.org/10.15405/EPSBS.2018.12.44

Gal, M. S., & Rubinfeld, D. L. (2019). Data Standardization. *New York University Law Review*, *94*. https://heinonline.org/HOL/Page?handle=hein.journals/nylr94&id=755&div=&collection=

Gartner. (2023, April). *Gartner Forecasts Worldwide Semiconductor Revenue to Decline 11% in 2023*. https://www.gartner.com/en/newsroom/press-releases/2023-04-26-gartner-forecasts-worldwide-semiconductor-revenue-to-decline-11-percent-in-2023

Gibson, C. (1982). Financial ratios in annual reports - ProQuest. *The CPA Journal (Pre-1986)*, *52*(9). https://www.proquest.com/openview/4a9fdd0afc3bbaf4e95ba06ced9cad37/1?pq-origsite=gscholar&cbl=41797

Goodall, R., Fandel, D., Allan, A., Landler, P., & Huff, H. (2002). *Long Term Productivity Mechanisms of the Semiconductor In-dustry*.

GVR. (2023). *Artificial Intelligence Market Size, Share & Trends Analysis Report By Solution (Hardware, Software, Services), By Technology (Deep Learning, Machine Learning, NLP), By Function, By End-use, By Region, And Segment Forecasts, 2023 - 2030*. https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-market

Heck, S., & Pinner, D. (2007). *Creating Value in the Semiconductor Industry 2*.

Helms, C. R. (2003). Semiconductor Technology & Manufacturing Status, Challenges, and Solutions — A New Paradigm in the Making. *Characterization and Metrology for ULSI Technology*, *683*, 63–73. https://doi.org/10.1063/1.1622453

Higgins, R. C. (2009). *Analysis for financial management*. 448. https://blackwells.co.uk/bookshop/product/Analysis-for-Financial-Management-by-Robert-C-Higgins-Jennifer-L-Koski-Todd-Mitton/9781265042639

Ho Yeap, K., & Javier Sayago Hoyos, J. (2020). *Integrated Circuits/Microchips*.

Hristozov, Y. (2021). Financial Analysis as a Key Component of the Business Valuation. *International Business and Accounting Research Journal*, *5*(1), 9–21. https://doi.org/10.35474/ibarj.v5i1.158

Hu, H. (2024). Navigating The Integrated Circuit Industry: Definitions, Evolution, Innovations, And Challenges. *Highlights in Science, Engineering and Technology*, *81*, 226–231. https://doi.org/10.54097/fb9z0110

Intel. (2014). *Intel Discloses Newest Microarchitecture and 14 Nanometer Manufacturing Process Technical Details*. https://www.intel.com/content/www/us/en/newsroom/home.html

Internation Roadmap for Devices and Systems™. (2020). *Future of Semiconductor Performance - IEEE IRDS™*. https://irds.ieee.org/topics/future-of-semiconductor-performance

Irwin, D. A., & Klenow, P. J. (1994). *Learning-by-Doing Spillovers in the Semiconductor Industry*. https://www.jstor.org/stable/2138784

Ivanova, R. (2019). *ABOUT THE ANALYSIS OF THE RATE OF RETURN OF ENTERPRISE'S ASSETS AND EQUITY*. https://ikm.mk/ojs/index.php/kij/article/view/1953/1952

Jewell, J., & Mankind, J. A. (2012). *What is Your ROA? An Investigation of the Many Formulas for Calculating Return on Assets*. https://www.researchgate.net/publication/254925117_What_is_Your_ROA_An_Investigation_of_the_Many_Formulas_for_Calculating_Return_on_Assets

Jory, S., Ngo, T., & Wang, H. (2021). Non-operating earnings and firm risk*. *Review of Financial Economics*, *39*(1), 95–123. https://doi.org/10.1002/RFE.1111

Kaiser, J. (2014). Dealing with Missing Values in Data. *Journal of Systems Integration*, 42–51. https://doi.org/10.20470/JSI.V5I1.178

Kozmetsky, G., & Yue, P. (1998). Comparative Performance of Global SC companies. *Omega*, *26*(2), 153–175. https://doi.org/10.1016/S0305-0483(97)00011-X

KPMG, & GSA. (2024). *Semiconductor outlook buoyed by AI and automotive*. https://kpmg.com/us/en/articles/2024/global-semiconductor-industry-outlook.html

KPMG India. (2024). *Semiconductor: The new driver of geopolitical balance* . https://kpmg.com/in/en/blogs/home/posts/2024/01/semiconductor-new-driver-of-geopolitical-balance.html

Larocca, R. (2005). Reconciling conflicting Gauss-Markov conditions in the classical linear regression model. *Political Analysis*, *13*(2), 188–207. https://doi.org/10.1093/PAN/MPI011

Li, Y. T. (2010). Best practice: An optimization of assets productivity in semiconductor manufacturing. *5th IEEE International Conference on Management of Innovation and Technology, ICMIT2010*, 880–884. https://doi.org/10.1109/ICMIT.2010.5492807

Lin, C.-S., & Huang, C.-P. (2011). Measuring competitive advantage with an asset-light valuation model. *African Journal of Business Management*, *5*(13), 5100–5108. https://doi.org/10.5897/AJBM10.681

Liu, H., Zhang, G., Zheng, X., Chen, F., & Duan, H. (2020). Emerging miniaturized energy storage devices for microsystem applications: from design to integration. *International Journal of Extreme Manufacturing*, *2*(4), 042001. https://doi.org/10.1088/2631-7990/abba12

Lyubimtsev, O. V., & Lyubimtseva, O. L. (2016). *Passage translated and paraphrased by the writer from Russian into English: Linear regression models in econometrics*.

Matiş, D., Vladu, A. B., & Negrea, L. (2010). Jones, Dechow And Kasznik Models Significance In The Romanian Economic Environment. *Annales Universitatis Apulensis Series Oeconomica*, *1*(12), 253–266. https://doi.org/10.29302/OECONOMICA.2010.12.1.25

McKight, P. E., & Najab, J. (2010). Kruskal-Wallis Test. *The Corsini Encyclopedia of Psychology*, 1–1. https://doi.org/10.1002/9780470479216.CORPSY0491

McKinsey & Company. (2024). *Exploring new regions: The greenfield opportunity in semiconductors*. https://www.mckinsey.com/industries/semiconductors/our-insights/exploring-new-regions-the-greenfield-opportunity-in-semiconductors

Mendez-Morales, A., Anzola-Morales, C., Ruiz-Acosta, L. E., & Camargo-Mayorga, D. A. (2024). Intangible assets and their effects on business performance: an analysis for Colombian companies. *Revista Galega de Economía*, *33*(1), 1–28. https://doi.org/10.15304/rge.33.1.9138

Microsemi PPG. (n.d.). *Gallium Nitride (GaN) versus Silicon Carbide (SiC) In The High Frequency (RF) and Power Switching Applications*. Retrieved March 23, 2024, from https://shop.richardsonrfpd.com/docs/rfpd/Microsemi-A-Comparison-of-Gallium-Nitride-Versus-Silicon-Carbide.pdf

MMR. (2023). *Artificial Intelligence Chip Market: Global Industry Analysis and Forecast (2024-2030) Trends, Statistics, Dynamics, Segment Analysis*.

https://www.maximizemarketresearch.com/market-report/artificial-intelligence-chip-market/185676/

Mordor Intelligence. (2023). *Automotive Semiconductor Market Share & Companies - Industry Report*. https://www.mordorintelligence.com/industry-reports/automotive-semiconductor-market

Msuha, B. (2019). Hypothesis Testing for the Association Between Categorical Variables: Empirical Application of Chi-square Test. *Hypothesis Testing for the Association Between Categorical Variables: Empirical Application of Chi Square Test*. https://doi.org/10.7176/MTM/9-2-02

Neff, E. (2023). *A Superatomic Semiconductor Sets a Speed Record | Columbia Quantum Initiative*. https://quantum.columbia.edu/news/superatomic-semiconductor-sets-speed-record

Nithin, K. S. (2023). A STUDY ON RETURN. ON EQUITY & RETURN ON ASSETS AT FEATHERLITE PRODUCT PRIVATE LIMITED WITH RESPECT TO OFFICE EQUIPMENTS, RAJANUKUNTE. *INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*, *07*(10), 1–11. https://doi.org/10.55041/IJSREM26666

Nordstokke, D., & Colp, S. M. (2014). Investigating the robustness of the nonparametric Levene test with more than two groups. *Psicologica*. https://www.semanticscholar.org/paper/Investigating-the-robustness-of-the-nonparametric-Nordstokke-Colp/a31df8468622260128edc7c33b108933696b2f8f

Oliver, R. W. (2001). Real Time Srategy. *Journal of Business Strategy*, *22*(4), 7–10. https://doi.org/10.1108/EB040177

Pandis, N. (2016). The chi-square test. *American Journal of Orthodontics and Dentofacial Orthopedics : Official Publication of the American Association of Orthodontists, Its Constituent Societies, and the American Board of Orthodontics*, *150*(5), 898–899. https://doi.org/10.1016/J.AJODO.2016.08.009

Park, J. H., Chung, H., Kim, K. H., Kim, J. J., & Lee, C. (2021). The Impact of Technological Capability on Financial Performance in the Semiconductor Industry. *Sustainability*, *13*(2), 1–20. https://doi.org/10.3390/SU13020489

Plackett, R. L. (1983). Karl Pearson and the Chi-Squared Test. *International Statistical Review / Revue Internationale de Statistique*, *51*(1), 59. https://doi.org/10.2307/1402731

Qolbi, F. A., Karisma, D. P., & Rosyadi, I. (2020). Macro Variable Effect Analysis and Non-Performing Financing (NPF) Against the Return On Asset (ROA) Islamic Banks In Indonesia Year 2008-2017. *Journal of Islamic Economic Laws*, *3*(1), 32–47. https://doi.org/10.23917/JISEL.V3I1.10170

R Core Team. (2021). *R: A Language and environment for statistical computing* (4.1).

Ram, M., & Chouhan, R. K. (2020). DUPONT ANALYSIS – A TOOL OF FINANCIAL PERFORMANCE ANALYSIS. *Indian Journal of Business Administration (IJBA)*, *13*(13), 07–12. https://www.busadmjnvu.org/JOURNAL%20IJBA%20DECEMBER%202020.pdf#page=19

Ross, S. A., Westerfield, R. W., & Jordan, B. D. (2022). *Fundamentals of corporate finance*. https://archive.org/details/fundamentalsofco0000ross_i4l4

Rutkowska-Ziarko, A. (2015a). The Influence of Profitability Ratios and Company Size on Profitability and Investment Risk in the Capital Market. *Folia Oeconomica Stetinensia*, *15*(1), 151–161. https://doi.org/10.1515/FOLI-2015-0025

Rutkowska-Ziarko, A. (2015b). The Influence of Profitability Ratios and Company Size on Profitability and Investment Risk in the Capital Market. *Folia Oeconomica Stetinensia*, *15*(1), 151–161. https://doi.org/10.1515/FOLI-2015-0025

SC-IQ: Semiconductor Intelligence. (2023, June). *Semiconductor CapEx down in 2023*. https://www.semiconductorintelligence.com/semiconductor-capex-down-in-2023/
Semiconductor Industry Association. (2024, February). *Global Semiconductor Sales Decrease 8.2% in 2023; Market Rebounds Late in Year*. https://www.semiconductors.org/global-semiconductor-sales-decrease-8-2-in-2023-market-rebounds-late-in-year/
Shaker Sultan, A. (2014). Financial Statements Analysis-Measurement of Performance and Profitability: Applied Study of Baghdad Soft-Drink Industry. *Research Journal of Finance and Accounting Www.Iiste.Org ISSN*, *5*(4). www.iiste.org
Shapiro, S. S., & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, *52*(3/4), 591. https://doi.org/10.2307/2333709
Shilov, A. (2024a, March). *Intel to Receive $8.5B in CHIPS Act Funding & Further Loans To Build US Fabs*. AnandTech. https://www.anandtech.com/show/21314/intel-to-get-85-billion-from-us-govt-to-build-fabs-in-the-us
Shilov, A. (2024b, March). *Micron Sells Out Entire HBM3E Supply for 2024, Most of 2025*. AnandTech. https://www.anandtech.com/show/21319/micron-sells-out-entire-hbm3e-supply-for-2024-most-of-2025
Shin, N., Kraemer, K. L., & Dedrick, J. (2017). R&D and firm performance in the semiconductor industry. *Industry and Innovation*, *24*(3), 280–297. https://doi.org/10.1080/13662716.2016.1224708
Singh, R., Chaudhary, P., & Gupta, C. P. (2023). *Defining Return on Assets (ROA) in empirical corporate finance research: a critical review*. https://doi.org/10.5281/zenodo.10901886
Singh, R., Gupta, C. P., & Chaudhary, P. (2023). Defining Return on Assets (ROA) in empirical corporate finance research: a critical review. *Empirical Economics Letters*, *23*(1), 25–36. https://doi.org/https://doi.org/10.5281/zenodo.10901886
Stewart, D., Hamling, D., Bucaille, A., & Crossan, G. (2021). *Semiconductor chip shortage | Deloitte Insights*. Deloitte Insights. https://www2.deloitte.com/uk/en/insights/industry/technology/technology-media-and-telecom-predictions/2022/semiconductor-chip-shortage.html
Studenmund, A. H. (2010). *Using Econometrics: A Practical Guide (5th Edition) (Addison-Wesley Series in Economics)*. 648. https://books.google.com/books/about/Using_Econometrics.html?id=TqtQmwEACAAJ
The jamovi project. (2022). *JAMOVI: open statistical software for the desktop and cloud* (2.3).
Tiron Tudor, A., Dima, Ş., Dima, B., & Valeria Raţiu, R. (2014). THE LINKAGE BETWEEN INTANGIBLES AND PROFITABILITY. *Annales Universitatis Apulensis Series Oeconomica*, *16*(1), 283–293.
Tsai, B.-H., & Chou, S.-B. (2015). Application of Multiple Output Data Envelopment Analysis in Interpreting Efficiency Improvement of Enterprise Resource Planning in Integrated Circuit Firms. *The Journal of Developing Areas*, *49*(1), 285–304. https://doi.org/10.1353/JDA.2015.0037
Tulyagankhodjaev, J. A., Shih, P., Yu, J., Russell, J. C., Chica, D. G., Reynoso, M. E., Su, H., Stenor, A. C., Roy, X., Berkelbach, T. C., & Delor, M. (2023). Room-temperature wavelike exciton transport in a van der Waals superatomic semiconductor. *Science*, *282*(6669), 438–442. https://doi.org/10.1126/SCIENCE.ADF2698
Udoayang, J. O., Uwah, U. E., & Asuquo, A. I. (2020). *International Journal of Management and Humanities (IJMH) Optimizing Return on Assets through Investment in Property, Plant and*

Equipment: Evidence from Listed Nigerian Manufacturing Companies. https://doi.org/10.35940/ijmh.J0950.0641020
U.S. Securities and Exchange Commission. (2024). *SEC.gov* . https://www.sec.gov/search-filings
Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual. Scotts Valley, CA: CreateSpace* (3.11.5).
Van Zant, P. (2014). *Microchip Fabrication: A Practical Guide to Semiconductor Processing* (6th ed.). McGraw Hill LLC.
Veblen, T. (1908). Fisher's Capital and Income. *Political Science Quarterly*, *23*(1), 112–128. https://doi.org/10.2307/2140944
Wai-chung Yeung, H., Huang, S., & Xing, Y. (2023). Global Value Chain Development Report 2023: Resilient and sustainable GVCs in turbulent times. *WTO*. https://www.wto.org/english/res_e/booksp_e/07_gvc23_ch4_dev_report_e.pdf
Weber, C. (2002). Knowledge Transfer and the Limits to Profitability: An Empirical Study of Problem-Solving Practices in Semiconductor Manufacturing and Process Development. *IEEE TRANSACTIONS ON SEMICONDUCTOR MANUFACTURING*, *15*(4). https://doi.org/10.1109/TSM.2002.804877
White, K. J. (1992). The Durbin-Watson Test for Autocorrelation in Nonlinear Models. *The Review of Economics and Statistics*, *74*(2), 370. https://doi.org/10.2307/2109675
Wilcox, R. (2005). Trimming and Winsorization. *Encyclopedia of Biostatistics*. https://doi.org/10.1002/0470011815.B2A15165
Wooldridge, J. M. (2012). *Introductory Econometrics: A Modern Approach | Request PDF*. *20*. https://www.researchgate.net/publication/303197102_Introductory_Econometrics_A_Modern_Approach
Zhai, K. (2023). The changing landscape of semiconductor manufacturing: why the health sector should care. *Frontiers in Health Services*, *3*, 1198501. https://doi.org/10.3389/FRHS.2023.1198501
Zimny, A. (2022). The Impact of Asset Structure on a Company's Financial Results Stability. *Finanse i Prawo Finansowe*, *4*(36), 107–128. https://doi.org/10.18778/2391-6478.4.36.07