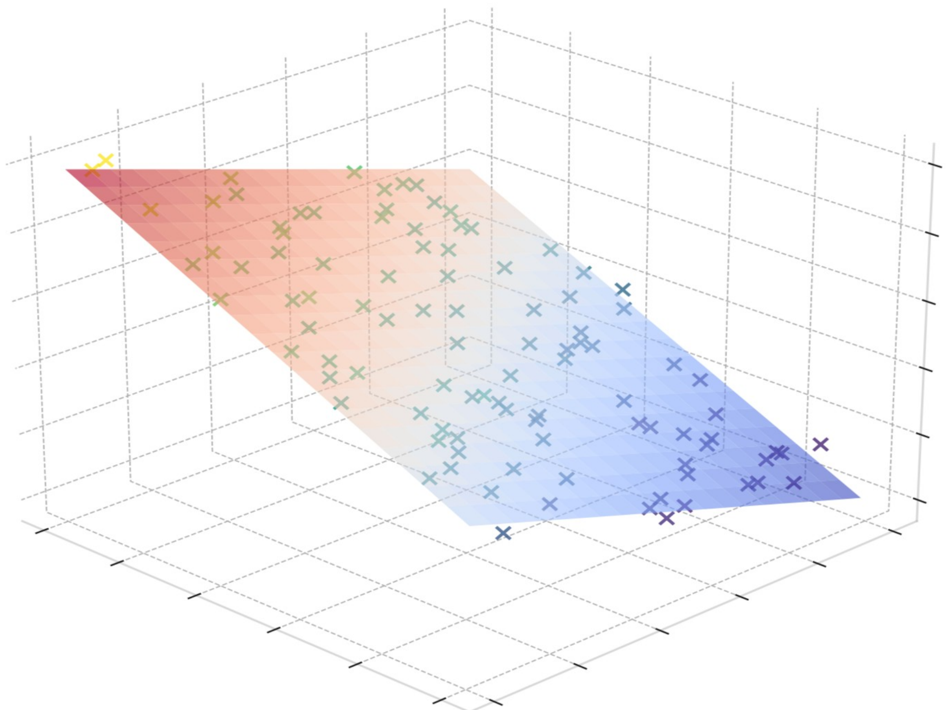
CHALMERS



Empirical Determinants of Enterprise Value in M&A Transactions

A Cross-Sector Regression Analysis of U.S. Small-Cap Public Company Deals (2005–2025)

Bachelors's thesis in Industrial Economy

JONATHAN EKELIND
PONTUS LUNDBORG
JACOB OLAUSSON
VIKTOR KAJANUS
HAMPUS LÖWING
ERIK TÖNDER

DEPARTMENT OF TECHNOLOGY MANAGEMENT AND ECONOMICS
DIVISION OF STRATEGY AND ENTREPRENEURSHIP

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2025
www.chalmers.se

Bachelor's Thesis TEKX18-25-13

Bachelors's thesis TEKX18-25-13

Empirical Determinants of Enterprise Value in M&A Transactions

A Cross-Sector Regression Analysis of U.S. Small-Cap Public Company Deals (2005–2025)

Empiriska faktorer som påverkar Enterprise Value vid M&A-transaktioner

En sektorsövergripande regressionsanalys av amerikanska börsnoterade small-cap-bolag (2005–2025)

JONATHAN EKELIND
VIKTOR KAJANUS
PONTUS LUNDBORG
HAMPUS LÖWING
JACOB OLAUSSON
ERIK TÖNDER

Department of Technology Management and Economics
Division of Strategy and Entrepreneurship
Chalmers University of Technology
Gothenburg, Sweden 2025

Empirical Determinants of Enterprise Value in M&A Transactions
A Cross-Sector Regression Analysis of U.S. Small-Cap Public Company Deals (2005–2025)

JONATHAN EKELIND VIKTOR KAJANUS
PONTUS LUNDBORG HAMPUS LÖWING
JACOB OLAUSSON ERIK TÖNDER

Supervisor: Christer Ljungwall, Technology Management and Economics
Examiner: Martin Löwstedt, Technology Management and Economics

Bachelor's Thesis TEKX18-25-13
Department of Technology Management and Economics
Division of Strategy and Entrepreneurship
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: 3D regression plot generated in Python.

Typeset in LaTeX
Gothenburg, Sweden 2025

Empirical Determinants of Enterprise Value in M&A Transactions
A Cross-Sector Regression Analysis of U.S. Small-Cap Public Company Deals (2005–2025)

JONATHAN EKELIND
VIKTOR KAJANUS
PONTUS LUNDBORG
HAMPUS LÖWING
JACOB OLAUSSON
ERIK TÖNDER

Department of Technology Management and Economics
Chalmers University of Technology

Abstract

Existing methods for calculating a firm's enterprise value (EV) inherently depend on subjective estimates and assumptions leading to variability based on the practitioner's judgment. This thesis uses a deductive, quantitative approach to investigate which factors influence EV in mergers and acquisitions (M&A). The analysis is based on secondary data from Capital IQ, covering U.S. small-cap public M&A transactions between 2005 and 2025. Using ordinary least squares (OLS) regression , the results show that market capitalization, net debt, and total assets are significant predictors of EV. These findings suggest that these financial variables play a key role in determining the value of a company in M&A deals. This data-driven valuation method offers a more objective alternative to traditional approaches.

Keywords: Enterprise Value, Mergers and Acquisition, Total Assets, Net Debt, Market Capitalization, Regression Analysis, Ordinary Least Squares.

Note: The thesis is written in English.

Sammanfattning

Befintliga metoder för att beräkna ett företags rörelsevärde (EV) bygger på subjektiva uppskattningar och antaganden vilket medför variationer beroende på den enskilda utövarens bedömning. Denna studie använder sig utav ett deduktiv och kvantitativ förförande i syfte att undersöka vilka faktorer som påverkar rörelsevärde inom fusioner och fövärv (M&A). Analysen genomfördes med hjälp av sekundärdata från databasen Capital IQ och avgränsades till publika small-cap transaktioner i USA mellan 2005 och 2025. En OLS-regression genomfördes i Python, där resultaten visade att börsvärde, nettoskuld och total tillgångar är signifikanta variabler för att förutsäga rörelsevärde. Detta resultat indikerar att dessa variabler spelar nyckerroller vid företagsvärdering inom M&A transaktioner. Denna datadrivna värderingsmetod erbjuder ett mer objektivt alternativ till traditionella tillvägagångssätt.

Nyckelord: Rörelsevärde, Fusioner och Förvärv, Totala tillgångar, Nettoskuld, Börsvärde, Regressionsanalys, Minsta kvadratmetoden.

Notera: Denna rapport är skriven på engelska.

Acknowledgements

We would like to express our gratitude to our supervisor, Christer Ljungwall, for his guidance, support, and helpful feedback throughout the construction of our regression analysis and the writing of this thesis. His advice has been invaluable during every stage of this project. We also want extend a thank you to Edvin Andersson, a former master student in finance, for the overview of the Capital IQ database he gave.

Gothenburg, May 2025
Jonathan Ekelind
Viktor Kajanus
Pontus Lundborg
Hampus Löwing
Jacob Olausson
Erik Tönder

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

AI	Artificial Intelligence
AIC	Akaike Information Criterion
B-P Test	Breusch-Pagan Test
CCA	Comparable Companies Analysis
CFNAI	Chicago Fed National Activity Index
COGS	Cost of Goods Sold
DCF	Discounted Cash Flow
EBIT	Earnings Before Interest and Taxes
EBITDA	Earnings Before Interest, Taxes, Depreciation, and Amortization
EBITDA-EBIT	EBITDA minus EBIT
EV	Enterprise Value
EV/EBIT	Enterprise Value to EBIT multiple
EV/EBITDA	Enterprise Value to EBITDA multiple
EV/Revenue	Enterprise Value to Revenue multiple
IEV	Implied Enterprise Value
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
M&A	Mergers and Acquisitions
Market Cap	Market Capitalization
NaN	Not a Number
NDA	Non-disclosure agreements
NPV	Net Present Value
OLS	Ordinary Least Squares
Revenue minus EBITDA	Revenue-EBITDA
P/BV	Price-to-Book Value Ratio
P/E	Price-to-Earnings Ratio
P/S	Price-to-Sales Ratio
QQ-plot	Quantile-quantile-plot
VIF	Variance Inflation Factor
WACC	Weighted Average Cost of Capital

Contents

List of Figures

List of Tables

1
Introduction

1.1 Background

Enterprise value (EV), as pointed out by Damodaran (2012), is an important concept in corporate finance and is often described as the theoretical takeover price of a firm. This makes EV an important financial metric for investors, analysts, and decision makers, providing a comprehensive measure of firm value that is particularly important in contexts such as mergers and acquisitions (M&A).

Damodaran (2012) continues with explaining that in M&A transactions, the EV is vital for deal structuring, negotiation, and strategic evaluation. The valuation of EV can influence important decisions related to pricing, financing, and expected returns. Traditional valuation techniques of EV include methods such as discounted cash flow (DCF) analysis, and relative valuation based on multiples, such as EV/EBITDA and EV/Revenue (Damodaran, 2012). These valuation techniques offer established and commonly used frameworks for estimating EV, either by projecting future cash flows or by benchmarking against comparable firms often based on size and industry.

The increasing availability of financial- and transaction-related data has enabled modeling of the relationships between firm characteristics and EV through statistical approaches such as regression analysis. These new data-driven approaches offer new opportunities to move beyond the standard methods previously mentioned in calculating EV (Wooldridge, 2013).

Considering how crucial EV is in corporate finance and its significance in M&A deals, a deeper empirical understanding of the financial and sector-specific determinants of EV remains a critical topic for both academic research and practical application.

1.2 Problem Analysis

EV is the standard measure of deal price in M&A, reflecting the total economic claim transferred from one party to another. However, as Blackham (2023) notes, when used by practitioners, valuation has largely been based on either intrinsic DCF models or simple multiples. While practically used and theoretically sound, Blackham (2023) highlights that there exists shortcomings within real transactions for these methods. When compiling a DCF model, long-term forecasts of cash flow and discount rate are required, making valuations highly sensitive to subjective inputs.

Minor adjustments in either the growth rate or the discount rate can significantly alter the estimated value, resulting in a wide range of outcomes that undermine reliability when projecting uncertain futures.

Multiple-based methods, however, depend on observable comparable companies, making them less sensitive to model inputs and quicker to implement, but they naturally carry their own biases. By standardizing prices into a single ratio, static multiples assume comparability across firms and ignore unique factors like differing capital structure, risk profiles or growth prospects. As Damodaran (2005) notes, the ease of using multiples can also be its weakness. A multiple can generate unreliable valuations when key drivers, such as risk and cash flows, vary significantly between companies. Moreover, market-based multiples are influenced by prevailing investor sentiment and market conditions, meaning if comparable companies are mispriced, the derived valuation will reflect that distortion. In summary, DCF and multiple-based methods simplify valuation, either through strong model assumptions or by applying a simple multiple, and as noted by Damodaran (2005), both approaches can be manipulated through choice of inputs or comparable companies.

This strong dependence on assumption driven models highlights a clear research gap: surprisingly little empirical work has directly examined transaction EV using econometric methods. Damodaran (2005, p. 3) observes that, despite valuations importance, “research into valuation models and metrics in finance is surprisingly spotty”. In particular, few studies have employed multivariate regression on deal data to determine the key drivers of EV. Some practitioners (e.g. Kaplan and Ruback, 1995) study the accuracy of EV/EBITDA as a predictor, but even these do not fit a multi-factor model of EV itself. One notable exception is Elmberger and Makdisi-Somi (2016), who built a multiple regression model for EV for the chemical sector. Their findings showed that regression models using financial independent variables marginally outperformed simple EV/EBIT multiples. However, that analysis was confined to one industry and even its authors note practical hurdles, such as complexity and data limitations, to broader adoption. As far as the literature reveals, no comprehensive cross-industry study has so far estimated EV as a dependent variable using transaction-level data. Instead, the literature remains focused either on equity returns, takeover premium, or static multiples benchmarks, leaving a gap in directly modeling the EV.

This gap suggests the need for a more statistically grounded, data-driven approach. M&A transaction values are inherently multi dimensional because they capture factors, such as the target company’s profitability, growth potential, and synergies, all influenced by current market dynamics. A simple ratio cannot capture this complexity. The selection of valuation multiples is closely tied to industry context, as comparable companies are often chosen from the same sector to ensure comparability in risk, capital structure, and profitability (Schueler, 2020). A regression framework can incorporate multiple predictors simultaneously, allowing factors like firm size and total assets to influence EV in a more flexible way. An empirical model can identify the most impactful variables practice and measure their affects statistically, instead of relying on subjective judgment or arbitrary assumptions.

To summarize, while the use of DCF and multiple based methods is commonly used among practitioners, their shortcomings are well documented (Blackham, 2023). Without econometric research on EV, valuations lacking a stringent, data-based multi-factor model, stay vulnerable to input biases and simplistic assumptions. Addressing this gap, by building and testing a cross-sector regression model for EV, would provide a rigorous benchmark for deal pricing and advance valuation theory in a M&A context.

1.3 Research Questions

Which financial variables significantly influence EV in M&A transactions?

How can these variables be integrated into a regression model to predict EV?

1.4 Purpose

This thesis aims to empirically examine the financial factors that drive EV in M&A transactions involving publicly traded small-cap companies in the U.S. By using a cross-industry, multivariate regression framework, the thesis seeks to identify which variables most significantly explain variations in transaction EV, thereby contributing to a more rigorous, data-driven foundation for valuation practice.

1.5 Limitations

This thesis is subject to some limitations arising from both the dataset selection criteria and the scope of available variables. These constraints were necessary to ensure analytical clarity and comparability. To maintain consistency and focus, specific inclusion criteria for the dataset were applied:

- **Implied Enterprise Value ($300M–$2B):** To ensure comparability and relevance, the analysis focused on transactions with an implied enterprise value between $300 million and $2 billion. This range captures small-cap companies, excluding both micro-cap transactions which tend to exhibit higher volatility, and mega-deals which frequently involve passive or minority stakes rather than central-oriented acquisitions. This approach aligns with the broader argument in Schoenmaker and Schramade (2023) that deal size impacts valuations, risks, and strategic relevance in M&A activity.

- **Geographic Focus (U.S.):** To ensure consistency in regulatory, legal, and economic framework, transactions are limited to U.S.-based targets. This approach minimizes valuation differences stemming from cross-border complexities, aligning with Schoenmaker and Schramade (2023) emphasis on the risks of large-scale M&A misvaluation in heterogeneous environments.

- **Transaction Type:** Only M&A were included, excluding other deal types such as divestitures, carve-outs, or joint ventures. This aligns with the paper's focus on M&A.

- **Percent Sought (>10%):** This threshold is motivated by both legal and economic reasons. Under U.S. securities law, ownership above 10% designates the acquirer as an "insider" subject to stricter regulation (U.S. Securities and Exchange Commission, 2024). These legal requirements may result in higher quality data being available for such transactions because insiders must adhere to more rigorous reporting standards. From an economic perspective, prior research highlights that substantial minority acquisitions, typically above 5%, are often strategically motivated and can give the acquirer influence over the target's operations or governance without obtaining full control (Ouimet, 2013). By focusing on transactions above 10%, deals of more strategic nature is captured.

- **Time Horizon (2005–2025):** The selection of M&A data from 2005 to 2025 is motivated by several reasons. Firstly, databases such as Capital IQ provides a higher data quality for data from 2000 and onward. While Barnes et al. (2014) describes improved data quality post-2000 in a Capital IQ competitor's database (Securities Data Company), similar improvements are assumed to have occured in CapitalIQ. Secondly, this time frame follows the introduction of the Sarbanes-Oxley Act of 2002, which increased corporate disclosure and internal control requirements for U.S. public companies, thereby contributing to more reliable financial reporting (Kenton, 2024).

2

Financial and Economic Theory

This section outlines the financial and economic concepts that form the foundation of this thesis. It begins by defining EV and related financial metrics, which are important for understanding the drivers of EV in M&A. The section also presents theoretical valuation frameworks, including DCF analysis, valuation multiples, and Comparable Company Analysis (CCA), explaining their relevance and application. Lastly, it introduces the Chicago Fed National Activity Index (CFNAI) as a macroeconomic indicator used to control for broader economic conditions during the valuation process.

2.1 Enterprise Value

EV is a widely used measure that represents the total value of a company, making it particularly relevant in the context of M&A. Unlike market capitalization, which only reflects the equity value of a company, EV accounts for the entire capital structure by including short- and long-term debt while subtracting cash and cash equivalents. This allows analysts and investors to evaluate a company's worth as if it were to be acquired, offering a more comprehensive view than equity market value alone (Fernando, 2024). Because EV uses figures from both the market and the company's financial statements, it better reflects the cost an acquirer would bear to purchase the business outright and assume its debts, minus any immediately accessible liquidity. The formula for calculating EV is presented in equation 2.1.

$$\text{EV} = \text{Market Cap} + \text{Net Debt} + \text{Preferred Equity} + \text{Minority Interest} \tag{2.1}$$

In equation 2.1, market capitilization is calculated by multiplying the company's current stock price by the total number of outstanding shares. Net debt includes total debt and cash, while preferred equity and minority interest are added if applicable. EV serves as the foundation for key financial ratios such as EV/EBITDA and EV/Sales, which are frequently used for relative valuation within and across industries. These metrics allow practitioners to compare firms more effectively, especially when differences in capital structure make traditional ratios like price-to-earnings (P/E) less meaningful (Fernando, 2024).

2.2 Implied Enterprise Value

In this thesis, IEV is obtained directly from the S&P Capital IQ platform. The value is not manually calculated but is instead based on figures provided by Capital IQ. IEV is calculated as the sum of Implied Equity Value and Net Debt Assumed, as shown in Equation 2.2.

$$\text{IEV} = \text{Implied Equity Value} + \text{Net Debt Assumed} \tag{2.2}$$

IEV reflects the total consideration to shareholders, adjusted for the ownership stake acquired, and is defined by Equation 2.3.

$$\text{Implied Equity Value} = \frac{\text{Common Shares Acquired} \times \text{Deal Value per Share}}{\text{Percent Acquired}} \tag{2.3}$$

Net debt assumed includes the company's net debt, preferred equity, and minority interest, as described in Equation 2.4.

$$\text{Net Debt Assumed} = \text{Net Debt} + \text{Preferred Equity} + \text{Minority Interest} \tag{2.4}$$

This methodology is detailed in S&P Global Market Intelligence (2021). All values are sourced from Capital IQ's transaction dataset and used without modification.

2.3 Net Debt

Net Debt is explained by MacDiarmid et al. (2018) as the company's total debt minus cash and cash equivalents. Thus, it is a financial metric that provides a clear picture of a company's financial health. It shows the debt burden after accounting for liquid assets that could be used to pay off debt. Net debt is calculated using Equation 2.5.

$$\text{Net Debt} = \text{Total Debt} - \text{Cash and Cash Equivalents} \tag{2.5}$$

MacDiarmid et al. (2018) continue by explaining that factors involving net debt are sometimes strong drivers of EV, while at other times they are less significant. Building on this, both Dang et al. (2019) and Hermuningsih (2013) explain that net debt affects EV both directly and indirectly. For example, Dang et al. (2019) point out that net debt is a key component in calculating EV, as it is directly included in the formula. Hermuningsih (2013) explains that net debt indirectly influences EV through the trade-off theory, which states that the value of a company increases with its debt up to a certain point. Thus, it appears that net debt plays an important role in shaping the actual EV.

Likewise, Akhtar et al. (2016) point out other relevant correlations between debt and a company's value. In their research, conducted through regression analysis, they found a significant relationship between debt and a company's value. They also mention that there is a difference between high- and low-growth companies in terms of how much debt influences the company's value. In summary, it can be concluded that debt must be optimized for a company to remain stable and attractive to investors. Thus, net debt is an important factor when determining the EV of a company.

2.4 Cost of Goods Sold and Operating Expenses

Shen and Fan (2017) explain that effective cost management is an important factor that influences the outcome of EV. They point out that high costs, especially those related to cost of goods sold (COGS) and operating expenses, can negatively affect a company's performance and, as a result, lower its EV. Furthermore, Shen and Fan (2017) emphasize that reducing these costs can lead to improved performance and consequently, a higher EV. Managing costs carefully allows companies to increase their profitability, and strengthen their financial position. Therefore, it can be concluded that both cost of goods sold and operating expenses play a significant role in shaping the outcome of a firm's EV.

2.5 Amortization and Depreciation

As explained by Zheng et al. (2023), depreciation and amortization are non-cash expenses included in the income statement. These represent the allocation of the cost of both tangible and intangible assets over their useful lives. Depreciation is associated with physical assets such as machinery or buildings, while amortization relates to expenditures for intangible assets such as patents or trademarks. In the income statement, depreciation and amortization reduce taxable income, whereas in the cash flow statement they bear no direct impact. Thus, without any actual outflow of cash, the company's taxable income decreases through the recognition of depreciation and amortization expenses. Zheng et al. (2023) note that this reduction in taxable income without affecting cash flow can enhance a company's EV. However, Zheng et al. (2023) also state that when EV is calculated using free cash flow, it becomes independent of depreciation and amortization. Nevertheless, the method chosen for calculating EV can influence how significant depreciation and amortization appear to be in relation to the EV.

2.6 Total Assets

Larger firms, which have greater total assets, tend to have a higher EV according to Juárez (2018). This is intuitive, as total assets represent everything the company owns, including cash, inventory, property, and equipment, and thus reflect its overall size. Additionally, Dang et al. (2019) also highlight the importance of total assets, describing them as a key factor for EV, since larger firms with more assets typically benefit from economies of scale and risk diversification. This suggests that total assets should positively affect EV.

2.7 Net Income Margin

Dang et al. (2019) explain that profitability has a positive effect on EV which is statistically significant. Similarly, Suliyanto et al. (2019) highlight that profitability is relevant for EV, specifically when measured in terms of net income margin. The formula for calculating net income margin is presented in Equation 2.6.

$$\text{Net Income Margin} = \frac{\text{Net Income}}{\text{Revenue}} \tag{2.6}$$

Net income margin measures a company's ability to generate profit from its sales. As explained by Suliyanto et al. (2019), a higher net income margin shows a greater ability to keep costs down relative to revenue, which in turn has a positive impact on EV.

2.8 Market Capitalization

As previously stated, market capitalization is a part of the equation for EV, namely in Equation 2.1. Thus, the value of market capitalization directly affects the outcome of EV. This is explained by Kaiser and Snyder (2013), who state that market capitalization is a fundamental component of EV. Since market capitalization is included in the calculation of EV, any change in market capitalization will directly impact it. Kaiser and Snyder (2013) explain that market capitalization represents the total value of a company's outstanding shares — this calculation is presented in Equation 2.7.

$$\text{Market Cap} = \text{Number of Outstanding Shares} \times \text{Price per Share} \tag{2.7}$$

The market cap reflects what the market believes the company is worth. Thus, an increase in market capitalization is typically driven by a rise in the price per share, or less commonly, an increase in the number of shares, which in turn leads to a higher EV (Kaiser & Snyder, 2013).

2.9 CFNAI

Brave (2020) states that CFNAI is an abbreviation of "The Chicago Fed National Activity Index". CFNAI monthly summarizes economic growth in the U.S. It consists of 4 different categories of economic activity, which according to Brave (2020) are: production and income; employment, unemployment and hours; personal consumption and housing; and sales, orders, and inventories. CFNAI is commonly used as an indicator of the current business cycle in the U.S. Brave (2020) explains that CFNAI has since 1967 been around 95 percent accurate in identifying U.S. recessions.

2.10 Valuation Theory and Frameworks

Valuation theory concerns the principles and assumptions used to estimate the economic value of firms and assets. It underpins financial decision-making by providing a structured way to assess value based on expectations of future performance, adjusted for risk and time. As Damodaran (2012) writes, valuation is about determining what an asset is worth today given its anticipated ability to generate future economic benefits.

2.10.1 Discounted Cash Flow Analysis

DCF is a widespread valuation method, with the objective of estimating the present value of a company based on its future cash flows (Fernandez, 2007). The principle underlying a DCF is that money's value changes over time. To account for this, a discount rate is being applied, reflecting both the investment's risk and the opportunity cost of capital (Fernandez, 2007). A DCF method typically involves two steps, first projecting future cash flows and discounting them to their present value, and second calculating a terminal to account for cash flows beyond the forecast period. A DCF is often best suited to companies whose cash flows is predictable and stable, since the goal is to measure intrinsic value independently of current market sentiment.

It is important to understand that a DCF strongly depends on the underlying assumptions made by the one compiling the inputs. Accurate forecasting future revenues, costs, capital expenditures, and other key factors requires a thorough analysis and a solid understanding of both the company's operations and the broader macroeconomic environment. To account for potential errors, practitioners often supplement a DCF with scenario or sensitivity analysis to model different outcomes based on varying inputs. Some criticize DCF models for fostering overconfidence in long-term forecasts, but supporters argue that, when executed properly, the method remains one of the most reliable and practical tools in valuation (Fernandez, 2007). As such, DCF continues to be a comprehensive and valuable tool in investment analysis, corporate finance, and strategic decision-making.

2.10.2 Comparable Company Analysis

Comparable Company Analysis (CCA) is outlined by Chen (2020) as a valuation method that estimates the value of a business by benchmarking it against similar firms within the same industry and of comparable size. The fundamental assumption underlying CCA is that companies operating under similar conditions should be valued using comparable financial multiples. Damodaran (2012) highlights EV/Revenue, EV/EBITDA, EV/EBIT, P/E, P/S and P/BV as examples of valuation multiples used. These multiples relate a company's EV or equity value to a financial performance metric such as book value, operating profit or revenue. By analyzing these metrics across a selected peer group, a relative valuation range and whether a particular firm is under- or overvalued compared to its peers can be derived.

Damodaran (2012) further notes that the choice of multiple is often sector specific. For example, EV/Revenue is frequently used for early-stage or low-margin businesses where EBITDA may be negative, while the price-to-book (P/B) ratio is more appropriate for financial institutions. Analysts often complement intrinsic valuation models, such as DCF, with CCA to gain a more holistic view of a company's value. Furthermore, CCA can incorporate data from relevant recent M&A transactions.

Multiples are commonly used in both investment banking and corporate finance settings. Lino (2024) explains that valuation multiples have limitations as they are often backward-looking and can be influenced by short-term market conditions and one-time events, making them susceptible to market sentiment. Moreover, applying these multiples requires careful judgment in selecting truly comparable companies and understanding the companies underlying value drivers. For a robust analysis, multiples are often used in conjunction with intrinsic valuation methods such as DCF analysis to cross-validate results (Lino, 2024). When applied thoughtfully, multiples remain a practical and effective tool for valuations across different industries.

3

Mathematical Theory

This section provides an overview of the mathematical and theoretical frameworks relevant to this thesis. Ordinary Least Squares (OLS) regression analysis and its key assumptions, such as linearity, homoscedasticity, and normality of residuals, are described. Furthermore, it explains the diagnostic tools Z-score and Cook's Distance for identifying outliers. Lastly, it addresses techniques for validating assumptions and improving model robustness, including hypothesis testing and model selection.

3.1 Key Assumptions of OLS Regression

When performing an OLS regression, various assumptions regarding the data are used, i.e. about the distribution of the residuals and other key features. According to Sarstedt and Mooi (2018) it is critical that these assumptions are met to yield a valid and reliable result from the OLS regression. However, if they are not met, the data can be modified to meet these requirements. Below, the key assumptions mentioned by Sarstedt and Mooi (2018) are listed.

1. There are linear relationships between the independent variables and the dependent variable, which means that the dependent variable can be written as a linear function of the independent variables.

2. The expected mean error is zero, it is important to note that this does not mean that the estimated mean error is zero. This is called exogeneity.

3. Homoscedasticity, meaning that the variance of residuals is assumed to be constant.

4. The residuals in the regression model are independent, which means that there is no autocorrelation between the residuals.

5. The residuals of the regression model are normally distributed.

6. The data is void of any collinearity or multicollinearity. Collinearity means that there exists a high level of correlation between two independent variables, and naturally multicollinearity is defined by a high level of correlation between three or more independent variables.

3.2 Multiple Regression Analysis

James et al. (2013) describe that when examining a variable that is dependent on more than one predictor, multiple linear regression is preferable as each independent variable gets a separate slope coefficient. When calculating and describing multivariate linear regression the formula presented in Equation 3.1 is used.

$$Y = \beta_0 + \beta_1 X_1 + ... + \beta_n X_n + \epsilon \tag{3.1}$$

In Equation 3.1, Y represent the sought or dependent variable, it depends on both the independent variables X_n and the error term ϵ. Each independent variable X_n is multiplied by its corresponding slope coefficient β_n.

James et al. (2013) explain that the error term, often denoted by ϵ, is a variable that captures both the influence of variables that affect the dependent variable but have been omitted as well as inherent randomness. The error term, or residual, accounts for the difference between the predicted and actual value of the dependent variable. When regression analysis is conducted, the residuals are assumed to be normally distributed (as discussed in Section 3.1).

3.3 Ordinary Least Squares

Brockwell and Davis (2016) describes that OLS finds the line or hyperplane that best fits the observed data by minimizing the sum of squared differences between the observed and predicted values. That is, OLS is useful in estimating the parameters of the linear regression model. The basic form of a linear model is presented in Equation 3.2.

$$y = X\theta + \varepsilon \tag{3.2}$$

Here, y is the vector of observed values, X is the matrix of predictors, θ is the vector of unknown parameters, and ε is the error term. The goal of OLS is to find the parameter vector $\hat{\theta}$ that minimizes the sum of squared residuals which is described mathematically in Equation 3.3.

$$S(\theta) = \sum_{i=1}^{n} (y_i - x_i'\theta)^2 \tag{3.3}$$

This minimization leads to the so-called normal equations, given by Equation 3.4.

$$X^\top X\hat{\theta} = X^\top y \tag{3.4}$$

When $X^\top X$ is invertible, the OLS solution is calculated as in Equation 3.5.

$$\hat{\theta} = (X^\top X)^{-1} X^\top y \tag{3.5}$$

3.4 Outliers

Wooldridge (2013) explains that OLS minimizes the sum of squared residuals, meaning that larger residuals are given disproportionately greater weight in the model. As a result, OLS is sensitive to extreme values, since such values can significantly impact the estimated coefficients. To address this, outliers needs to be handled carefully to reduce their influence on the model. Chatterjee and Hadi (1986) emphasize the importance of diagnosing influence rather than simply removing data without proper assessment. In other words, the upper and lower quantiles should not be excluded blindly. Their reasoning is that not all outliers are necessarily problematic, some may have little impact on the model, while others can significantly affect the results even if their residuals are relatively small. Therefore, each observation's influence should be carefully evaluated before making any exclusions. Chatterjee and Hadi (1986) explain that this can be done using influence diagnostic tools, such as Cook's distance and Z-Score.

3.4.1 Z-score

To detect outliers in data, the Z-score method can be applied. Z-score is one of the most commonly used techniques for identifying outliers in a univariate dataset. As described by Barnett and Lewis (1994), the Z-score represents how many standard deviations an observation is from the mean of the distribution. The Z-score for an individual observation x is calculated in accordance with Equation 3.6, where x is the observed value, μ is the mean of the sample, and σ is the standard deviation.

$$z = \frac{x - \mu}{\sigma} \tag{3.6}$$

This standardization allows values from different distributions to be compared on the same scale. Barnett and Lewis (1994) notes that values with a Z-score greater than +2 or less than -2 can be considered outliers, following the common rule of thumb based on the properties of the normal distribution. Z-score is particularly useful as a first step in cleaning data before applying statistical models, and can be used to screen for unusually high or low values.

3.4.2 Cook's Distance

Another diagnostic tool is Cook's Distance, which can play an important role in identifying observations that have a strong influence on regression results. This method, first introduced by Cook (1977), helps assess how much the estimated regression coefficients would change if a specific data point were removed. It does this by combining how far an observation is from the regression line, its residual, with how unusual it is in terms of its predictor values, its leverage. The formal definition of Cook's Distance for an observation i is presented in Equation 3.7, where $\hat{\beta}_{(i)}$ refers to the estimated coefficients without observation i, the matrix X is the design matrix of predictor variables, p is the number of predictors in the model, and $\hat{\sigma}^2$ is the estimated error variance. In practice, it is often easier to use a simplified version as seen in Equation 3.8.

$$D_i = \frac{(\hat{\beta} - \hat{\beta}_{(i)})^\top X^\top X(\hat{\beta} - \hat{\beta}_{(i)})}{p \cdot \hat{\sigma}^2} \tag{3.7}$$

$$D_i = \frac{r_i^2}{p} \times \frac{h_{ii}}{(1 - h_{ii})^2} \tag{3.8}$$

In Equation 3.8, r_i is the studentized residual, and h_{ii} is the leverage value for that observation. A higher D_i means the observation has a greater impact on the model. As a general guideline, values greater than the threshold $\frac{4}{n}$, where n is the number of observations, can be considered influential and should be removed (Cook, 1977).

3.5 Hypothesis Testing

This chapter presents the foundation of hypothesis testing and provides the reader with the information needed to interpret the performance of the regression model. P-value and R^2 are introduced followed by two statistical tests called F-test and T-test.

3.5.1 P-value

James et al. (2013) explains that the p-value assesses whether there is a statistically significant relationship between a predictor and a response variable in a regression model. It is defined as the probability of getting a test statistic as extreme as the one observed, assuming that the null hypothesis is true. Mathematically, if t is the observed value of the test statistic, then the p-value is given by Equation 3.9.

$$\text{p-value} = \mathbb{P}(\text{Test statistic} \geq t \mid H_0 \text{ is true}) \tag{3.9}$$

Furthermore, James et al. (2013) notes that the p-value helps decide whether to reject the null hypothesis, which usually states that there is no relationship between the predictor and the response. A small p-value, typically less than 0.05, suggests that there is strong evidence against the null hypothesis, meaning that the predictor is likely related to the response. A large p-value means there is not enough evidence to say that there is a relationship.

3.5.2 R^2

The R^2 statistic assess how well a regression model fits the data. It is defined as the proportion of the variance in the response variable that is explained by the predictors. The mathematical expression is shown in Equation 3.10.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3.10}$$

The R^2 value lies between 0 and 1, where a value close to 1 indicates a strong model fit and a value near 0 suggests that the model fails to explain much of the variation in the data (James et al., 2013).

3.5.3 F-test

As presented by Sureiman and Mangera (2020), an F-test assesses whether a regression model provides a better fit to the data than a model with no independent variables. This means that the F-test can be used as an overall performance-measure for the regression model. However, the significance of the F-test must be taken into consideration. If the p-value is less than the chosen level of significance, there is evidence to suggest that the model performs significantly better than a model with no independent variables. It is also true that if none of the independent variables are statistically significant, the F-statistic is generally insignificant as well.

Sureiman and Mangera (2020) further explain that there are several underlying assumptions to the F-test which need to be met for the test to be valid. The relationship between the dependent and the independent variables should be linear. The differences between the observed values and the values predicted by the regression model should be approximately normally distributed. In other words the residuals should be normally distributed. Additionally, Sureiman and Mangera (2020) explain that the independent variables should not be highly correlated with each other which is measured in multicollinearity. The fourth underlying assumption mentioned by the authors is homoscedasticity.

3.5.4 T-test

Watts (2022) explains that if the F-test proves valid, at least one coefficient in the regression model differs from 0 and the overall performance of the model is accepted as adequate. However, the F-test does not specify how each independent variable is related to the dependent variable. For this, a t-test is used to examine the relationship between the dependent variable and each individual regression coefficient. Watts (2022) elaborates on the null hypothesis in a t-test and explains that the null hypothesis is that there is no relationship between the independent and the dependent variable. This means that the alternative hypothesis is that there is a relationship between the dependent and independent variable. The p-value of the t-test is then compared to the chosen significance level to determine if the null-hypothesis should be rejected. If the p-value is lower than the significance level, the null-hypothesis stating that there is no relationship between the regressions independent variables and the dependent variable is rejected. This means that a p-value lower than the significance level is equivalent to the independent variables having a significant impact on the model and vice versa.

3.6 Errors

In the following section, potential errors related to the OLS assumptions discussed in Section 3.1 will be more thoroughly described as well as what methods can be used to ensure that all assumptions are met.

3.6.1 Linearity

There exists no specific test to determine if there is linearity between the dependent variable and the independent ones. The best way to check for linearity is to make scatter plots according to Sarstedt and Mooi (2018). From the scatter plots, the linearity is rather easily dismissed or not by visually inspecting the plot. According to Sarstedt and Mooi (2018), linearity yields a better fit of the OLS model. However, if the relationship in fact is not linear, one might use some transformation, e.g. squaring the variable to make a better model fit. Furthermore, Sarstedt and Mooi (2018) states that the most important feature of linearity is the ability to express the dependent variable as a linear function of all the independent variables. This kind of linearity is somewhat difficult to test, but as Sarstedt and Mooi (2018) state if the OLS regression yields satisfactory test scores, it is highly plausible that the relationship is linear.

3.6.2 Autocorrelation

The assumption of autocorrelation handles the relationship between the residuals. If there exists any autocorrelation between the residuals, a change in one residual will result in a change in another residual according to Sarstedt and Mooi (2018). However, this is mostly an issue when working with a time-series rather than a dataset consisting of different transactions. Nevertheless, a test must always be performed to ensure that there is no autocorrelation between the residuals. Sarstedt and Mooi (2018) states that one test for this is the so-called Durbin-Watson test. In the Durbin-Watson test, the null hypothesis asserts that no autocorrelation exists. If the p-value is above 0.05, we fail to reject the null hypothesis, indicating that the assumption of no autocorrelation is fulfilled.

3.6.3 Endogeneity

If the expected values of the residuals are not equal to 0, there is endogeneity. Since exogeneity is an assumption in OLS it is important to verify its presence. However, there are no statistical tests to verify that the expected value of the residuals is zero.

Therefore, while exogeneity is assumed if evidence supports it after examining potential causes of endogeneity, some plausible reasons for endogeneity will be outlined. Lang (2014) lists some possible reasons for endogeneity, such as, sample selection bias, simultaneity, missing relevant independent variables, and measurement errors.

The different reasons for endogeneity are explained by Lang (2014). Sample selection bias means that some other criterion is affecting the probability of being a part of the sample. Simultaneity is best explained as the dependent variable affecting at least one of the independent ones. Missing relevant independent variables is rather self-explanatory and the same goes for measurement errors. Sarstedt and Mooi (2018) also state that endogeneity might indicate that relevant independent variables are missing.

3.6.4 Homoscedasticity

It might be that the variances of the residuals are not constant, this is instead called heteroscedasticity. This is not desired since it prohibits conclusions to be drawn. To test whether the residuals are homoscedastic there are several different methods, but the most frequently used one is the Breusch-Pagan test (B-P test). The B-P test checks if the residuals of the model are heteroscedastic (Breusch & Pagan, 1979). The B-P test consists of both a Lagrange multiplier test and a F-test, and the null hypothesis of the test is that the residuals are homoscedastic. If the p-value for both the Lagrange multiplier test and the F-test are above 0.05 the null hypothesis cannot be rejected, and the residuals are considered homoscedastic, otherwise the null hypothesis is rejected and the residuals are deemed heteroscedastic. According to Lang (2014), a commonly used method to avoid heteroscedasticity is to log-transform the dependent variable, i.e., taking the natural logarithm of the dependent variable.

3.6.5 Collinearity and Multicollinearity

Another common error with respect to the data is multicollinearity between the independent variables. This means, as previously mentioned, that there is some correlation between the independent variables. There are several tests to check whether multicollinearity exists, and the most common one is to check the Variance Inflation Factor (VIF) (Sarstedt & Mooi, 2018). The formula for VIF presented by Thompson et al. (2017), is shown in Equation 3.11.

$$VIF = \frac{1}{1 - R^2} \tag{3.11}$$

In Equation 3.11, R^2 is the multiple correlation coefficient. There are some generally accepted benchmarks for VIF which state that if the VIF is 10 or more there exists multicollinearity, otherwise multicollinearity does not exist (Sarstedt & Mooi, 2018). If some independent variables exhibit multicollinearity, one must simply remove one of them and check if the VIF-score improves. If the VIF-score improves the multicollinearity have been removed.

According to Westlund and Petrova (2018) a common error with regards to collinearity is the dummy variable trap. Dummy or binary variables are constructed when quantifying information that is of qualitative nature. If the criteria is met the variable takes the value 1, otherwise 0. When constructing three or more such variables, strong multicollinearity can occur, since one variable can be perfectly predicted using the other variables. To avoid this trap, one of the binary variables is omitted and will instead be used as the reference variable.

3.6.6 Normality

A regression analysis is based on the underlying assumption that the mean of the residuals from the regression are normally distributed. Since this is a crucial assumption, it is of interest to examine whether the residuals actually are normally distributed. As mentioned by Das and Resnick (2008), a relevant question to ask when handling data is whether the sample of data comes from a certain distribution. They further explain that a quantile-quantile-plot (QQ-plot) is a common way to test how well a certain distribution fits a sample of data. The QQ-plot compares the quantiles of a given sample data with the quantiles of a theoretical distribution. The x- and y-axis consist of the theoretical and sample quantiles respectively, and the values are plotted against a 45° straight line going across the plot. If the quantiles of the sample data perfectly match the quantiles of the theoretical distribution, all values will be on the straight line since they are equal.

According to Mishra et al. (2019), one of the most common ways to analyze if a data sample is normally distributed is using a histogram. The authors explain that the data can be assumed normally distributed if two conditions are met; the histogram resembles a bell curve, and is symmetric around its mean. To validate this, the theoretical normal distribution can be plotted as a reference and the actual data is included bart as part of a histogram. The bar height is determined by the number of observations close to that value. This way it is observable if the data resembles a bell curve and if it is symmetric around its mean.

3.7 Variable Selection

The following section explains important points to consider when choosing variables in multiple regression analysis. It includes the use of proxy variables for factors that cannot be directly measured, the role of control variables, how to handle variables that are not statistically significant, and how to use model selection tools like the Akaike Information Criterion (AIC).

3.7.1 Proxy-variables

In applied research, it is common to encounter important variables that cannot be directly measured. In these cases, proxy variables are used to represent the unobserved factors (Wooldridge, 2013). A proxy variable is expected to be related to the unobserved variable and not related to the error term. When chosen carefully, proxy variables can improve the model's accuracy. However, they should be used with caution, as poor proxies can create new problems. According to Wooldridge (2013), the decision to include a proxy should be based on theory rather than just statistical results. Including proxies can aid in building models that better reflect the real-world relationships.

3.7.2 Control Variables

According to Wooldridge (2013), certain variables are included in the multiple regression analysis not because they are of primary interest, but because they help isolate the effect of variables of interest. These are referred to as control variables and play a key role in achieving an, all else equal, interpretation of regression coefficients. As Wooldridge (2013) further explains, the purpose of a control variable is to account for other factors that influence the dependent variable, so that the estimated effect of the variable of interest is not biased by omitted variables.

3.7.3 Non-significant Variables

James et al. (2013) explains that in multiple linear regression modeling, an independent variable with a high p-value, that is, a p-value greater than 0.05, is considered statistically insignificant. This implies that it does not contribute meaningfully to explaining the variation in the dependent variable. Thus, if independent variables have large p-values, they do not contribute significantly to explaining the variation in the outcome, and their coefficients are likely close to zero. According to James et al. (2013), such variables could therefore be removed from the model.

When building a multiple linear regression model, some control variables may turn out to be statistically insignificant. In such cases, James et al. (2013) suggest that it is often reasonable to consider removing these variables from the model, especially if they lack theoretical support. However, before removing them, it is important to consider why the variables were included in the first place. Some control variables may still be useful for ensuring that the effects of other variables are measured accurately and can be included based on theoretical reasoning rather than on solely statistical test results. Through this reasoning James et al. (2013) describes that control variables can help increase the model's validity even though its individual effect is not statistically significant.

3.7.4 Akaike Information Criteria

According to Cavanaugh and Neath (2019), it is common practice to choose a criterion to evaluate whether a model has an appropriate structure when performing a regression analysis. This criterion determines whether the model executes an optimal balance between good fit and simplicity. A model with this kind of balance should be able to generalize well and describe or predict data from the same phenomenon. In an attempt to find a general model, the risk of ending up with a too simplistic model increases. Conversely, if the model is fitted too well generality is lost and the accuracy of predictions will decrease. In short, a selection criterion will rule out a model if it is either too simple or too overfitted.

Cavanaugh and Neath (2019) further explain that one of the most common model selection criteria is the Akaike Information Criterion (AIC). The AIC is derived from the principle of maximum likelihood and compares different models to each other when applied to the same data. The model with the lowest AIC-value is considered the superior one.

3.8 Prediction Error Metrics

In this section, two common prediction error metrics are described, these two are commonly used for evaluating the performance of a model.

3.8.1 Mean Absolute Error

The Mean Absolute Error (MAE) is a measure of forecast accuracy that calculates the average magnitude of the forecast errors, without considering their direction. MAE's mathematical definition is presented in Equation 3.12.

$$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n}|e_t| \tag{3.12}$$

Here $e_t = Y_t - F_t$ is the forecast error at time t (Hyndman & Koehler, 2006).

3.8.2 Mean Absolute Percentage Error

The Mean Absolute Percentage Error (MAPE) expresses forecast errors as a percentage of the actual values, providing a scale-independent measure. It is defined as in Equation 3.13.

$$\mathrm{MAPE} = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{e_t}{Y_t}\right| \tag{3.13}$$

In Equation 3.13 e_t is defined as in Section 3.8.1 and Y_t is the observed value at time t (Hyndman & Koehler, 2006).

4

Method

This section describes the methodological decisions made in this thesis. A deductive and quantitative approach was applied to test the factors influencing EV in M&A-transactions. To analyze the relationship between EV and its potential drivers, multiple regression analysis using OLS was used. The data used for the analysis was collected from the Capital IQ database and processed using Python-based tools. This section also addresses validity, reliability, and ethical use of data, followed by a method critique that reflects on the challenges and considerations involved in applying the selected methods and data sources.

4.1 Methodological Approach

This thesis used a deductive research approach, which Denscombe (2010) explains involves starting from general principles or theories and then testing these theories against real-world evidence. The deductive approach included hypothesis testing, in which theoretical ideas from previous research were used to design a thesis that examined whether these ideas held true in a specific setting. In this research, theories from financial and valuation theory formed the basis for identifying a set of factors believed to affect EV in M&A. These factors were converted into measurable variables and analyzed using statistical methods to assess their influence on the outcome. Denscombe (2010) stresses that a deductive approach is particularly suitable when the aim is to confirm or reject existing theories rather than to generate new ones. This made the deductive approach a suitable choice for this thesis. Denscombe (2010) also notes that this approach works well when there is data available for quantitative analysis.

To support this deductive design, the thesis used a quantitative research approach. In the context of M&A, identifying the factors that influence the determination of EV required a method capable of systematically analyzing performance ratios, financial metrics, and market indicators. Denscombe (2010) notes that when the goal is to measure variables numerically and uncover patterns or relationships among them, a quantitative approach is appropriate. Similarly, Creswell and Creswell (2018) emphasizes that quantitative research is well suited to examine relationships between measurable variables. Consequently, this thesis applied a quantitative approach, which enabled statistical testing of the factors impacting EV and helped determine which variables had the most significant influence on valuation outcomes.

To effectively capture the complexity of these relationships, multiple regression analysis was used. As Wooldridge (2013) explains, relying on a single independent variable to predict a dependent variable limits the ability to draw reliable conclusions, as it overlooks the influence of other relevant factors. Multiple regression analysis addresses this limitation by allowing several independent variables to be analyzed simultaneously, thus offering a more accurate and comprehensive understanding of what drives EV. Given that valuation is rarely determined by a single metric, this method is particularly well suited to the thesis objective of modeling EV with multiple influential independent variables in the context of M&A.

Furthermore, as Wooldridge (2013) explains, OLS is one of the most widely used methods for estimating parameters in multiple regression models. Similarly, Hayashi (2000) recognizes OLS regression as an important tool in econometrics. Given its usefulness in identifying relationships between multiple variables, OLS was considered an appropriate method for this thesis. In addition, Hayashi (2000) highlights the relevance of OLS in economic research, as it helps explain how different variables influence an outcome, confirming that this method is appropriate for the purposes of this thesis.

4.2 Data Source and Analytical Tools

Conducting primary data collection for quantitative research requires gathering a large amount of data, as quantitative research relies on extensive datasets. As a result, primary data collection can be both time-consuming and costly. Creswell and Creswell (2018) highlights that, for this reason, researchers often turn to secondary data sources such as large national databases and institutional records. Given the limited resources of this thesis, particularly in terms of funding, the quantitative component was conducted using secondary data collection. The data collected for the quantitative analysis was predominantly sourced from Capital IQ, a database developed by S&P Global. It was chosen on account of its broad applicability for secondary data collection in both academic and professional contexts. Its comprehensive and detailed coverage of financial data, particularly within transaction screenings, including detailed financial statement information at the transaction date, made it especially well-suited for this thesis.

After the initial screening in Capital IQ, the data was exported to Excel and then processed in Python. Several Python libraries were used; Pandas was used to organize and clean the data, while NumPy helped with basic calculations and data transformations. Statsmodels was used to run the OLS regression, check model assumptions, and perform diagnostic tests like the B-P test for heteroscedasticity. Seaborn and Matplotlib were used to create graphs showing correlations, model results, and residuals. Scikit-learn was used to split the data into training and testing sets and to measure the model's performance using metrics like MAE and R^2. Outliers were detected using the Z-score method from Scipy and Cook's Distance from

Statsmodels. To check for multicollinearity, the VIF was calculated. Together, these tools made it possible to prepare the data, build the regression model, and carefully test the model's performance.

4.3 Validity and Reliability

One crucial step in the data collection phase is to ensure that the data is credible and can be reliably trusted as accurate and true. Both Denscombe (2010) and Yin (2003) emphasize the importance of internal validity, external validity, and reliability in establishing credibility within a research text. More specifically, validity ensures that the data is accurate, precise, and appropriate for answering the research question. Reliability determines whether a research instrument produces consistent results across different occasions, while external validity assesses whether research findings can be applied to similar situations beyond the studied case. These factors were applied and carefully examined during the data collection and filtering of literature for this thesis.

4.4 Ethics

To ensure ethical use of data in a quantitative context, it must be clear where the data originates from and who was responsible for its initial collection according to Kara (2018). Consequently, this is done in Section 4.2. Furthermore, all policies and terms governing the respective databases was strictly followed. In addition, Kara (2018) emphasizes that accurate citation demonstrates respect by acknowledging the work of those referenced and giving them proper credit for their efforts. Furthermore, it also shows respect for readers by enabling the writing to be followed and understood, along with the foundation of the arguments presented. Similarly, Wallén (1996) explains the importance of acknowledging the writers referenced in an academic text, not only to avoid plagiarism, but also as an ethical practice that respects the hard work of others. Therefore, all information taken from sources will be properly referenced, ensuring that authors are acknowledged for their work each time it is used in the thesis.

4.5 Method Critique

Although the use of secondary data from Capital IQ was suitable for this thesis, it is important to acknowledge its limitations. As Creswell and Creswell (2018) point out, secondary data can be connected with challenges because the researcher has limited control over the relevance of the information for the specific thesis. In this case, the financial data from Capital IQ was originally collected for a wide range of purposes and hence is not collected specifically for the thesis purpose but rather for commercial purposes towards a wide range of consumers, both professionals and academicians. Hence, the data may not perfectly align with the specific research question of this thesis. Furthermore, Creswell and Creswell (2018) emphasize that secondary data is often gathered with different objectives in mind, which can lead to limitations when using it as a data source. Thus, although Capital IQ provides detailed and widely trusted financial information, these factors could still introduce biases when specifically analyzing EV in M&A.

As previously discussed, OLS is connected with the assumptions of homoscedasticity, expected mean error being zero, no autocorrelation between errors, residuals of the model being normally distributed, a linear relationship between independent and dependent variables, and the data being assumed to have no collinearity or multicollinearity. Regardless of the fact that the data was modified to meet these requirements, it should still be considered that this is a limitation of OLS. James et al. (2013) notes that OLS assumes perfect conditions, but these are rarely fully met in practice. If the assumptions are not met, it could lead to misleading errors and poor predictions. It is important to be aware of these limitations, especially in this case due to the complexity of financial data in M&A transactions.

As Creswell and Creswell (2018) explain an important limitation in research is how well the results can be generalized to other situations. In this thesis, the data was, as mentioned in section 1.5, limited to small-cap companies, transactions in the U.S., and deals where more than 10% of the company was bought. Because of the specific scope of the thesis, the results may not directly reflect conditions in contexts such as larger companies, other countries, or smaller transactions. Since financial markets and deal structures can differ across firm sizes and geographic regions, it is important to consider these factors when applying the findings beyond the studied group.

5

Data

This chapter describes what filters were used in the data screening process. Moreover, the approach to data handling is presented and explained. An overview of the dataset and its variables as well as a brief introduction to the analytical approach, are also provided.

5.1 Data Collection

The collection of data was done using S&P Capital IQ and its internal screening function. The Capital IQ screening function enables the selection of filters for data as well as data fields where variables of interest are highlighted. The screening was performed in the overarching category "transactions" , consequently included data points were limited to those derived from transactions. In addition, five filters were applied to the screening; these are presented below.

- Implied Enterprise Value ($USDmm, Historical rate): Between 300 and 2000
- Geographic Locations (Target/Issuer): The U.S. (Primary)
- Transaction Types: Mergers/Acquisitions
- Percent Sought (%): Greater than 10
- All Transactions Announced Date: 1/1/2005 - 3/27/2025

When conducting the screening, the transactions where limited to an IEV in the range of $300 million to $2000 million. The included transactions are also filtered geographically so that all target companies that are incorporated outside the U.S. are omitted. Percent sought is set to greater than 10% and all transactions included occurred between the 1st of January 2005 and the 27th of March 2025. The choice to include these specific filters are in line with what is presented in Section 1.5.

5.2 Independent Variables

The data fields chosen from S&P Capital IQ are included in the dataset and consist of the following variables: market capitalization, net debt, Revenue minus EBITDA (Revenue-EBITDA), total assets, net income margin, EBITDA minus EBIT (EBITDA-EBIT), and CFNAI. Market capitalization is included in the dataset as a control variable to take into account the different sizes of companies. The market's overall willingness to buy is taken into account using the control variable CFNAI. The final control variables that are included in the model are the main sector in which the target company operates. These where included to capture the inherent difference between industries and sectors. Additional variables are included in data fields because they can impact a company's EV in accordance with the economic theory in Section 2. The inclusion of Revenue-EBITDA is motivated by the fact that it can serve as a proxy for a company's operating costs and COGS which in turn can explain EV. This relationship simply stems from the income statement's nature. As discussed in Section 2.5, EV can be affected by depreciation and amortization. In some cases, it appears to influence EV through tax deductions, while in other cases, it seems to be irrelevant, making it an interesting variable to include as an independent variable. EBITDA-EBIT is therefore included as a proxy variable to represent a company's depreciation and amortization.

In addition to the mentioned variables, all sectors available in S&P Capital IQ were included in the dataset. In this case, sector is defined as the main sector in which the acquired company is operating within. These sectors were consumer discretionary, consumer staples, energy, financials, health care, industrials, information technology, materials, real estate, utilities. Since only one of these sectors is associated with each transaction, they are handled as binary variables. The correlation between the independent variables is presented in Figure 5.1.

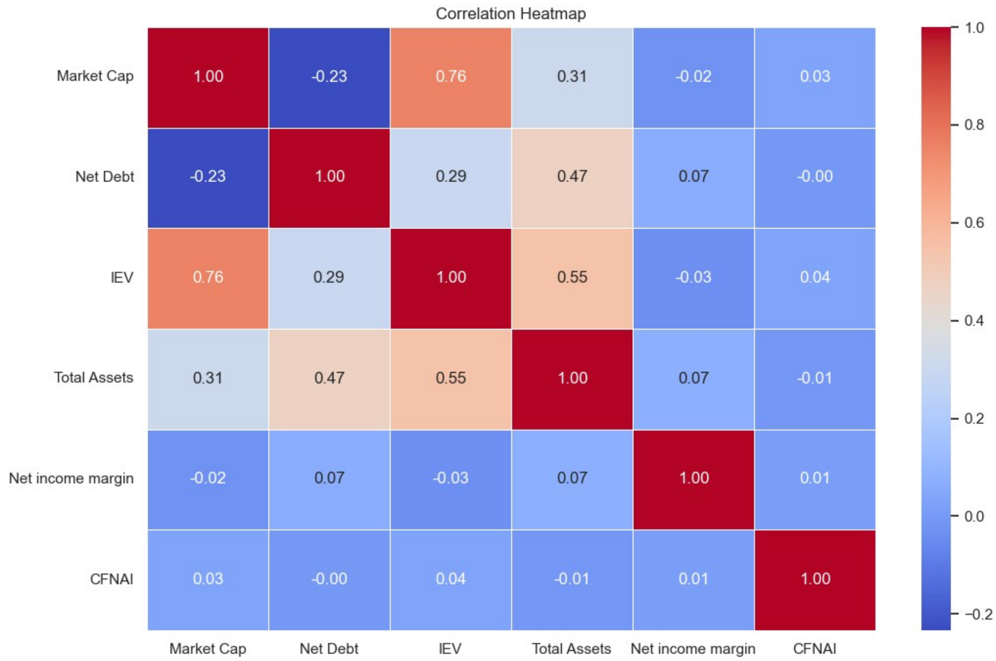


Figure 5.1: Correlation Matrix of Independent Variables

From Figure 5.1 it is noticeable that market capitalization and EV are highly correlated with a correlation coefficient equivalent to 0.76. As noted by Sarstedt and Mooi (2018), the correlation between the independent variables should be low. However, since this relationship is one between the dependent variable and one of the independent variables it is not affected by the multicollinearity assumption. Bahna (2009) further explain that a strong correlation between a dependent and independent variable is not only expected, but desired since there is an implied causal relationship between the two variables in a regression model. Subsequently, the high correlation between market capitalization and EV is not a cause of concern. Another noticeably high correlation is the one between EV and total assets which again, is not a cause of concern following what was previously stated by Bahna (2009). Net debt and total assets however, are two independent variables and should therefore not be highly correlated. As shown in Figure 5.1, these two variables have a correlation-coefficient of 0.47. Bahna (2009) states that the correlation is considered weak if this value is less than or equal to 0.4 and considered moderate if the value lies between 0.4 and 0.8. Since the correlation between net debt and total assets is shown to be 0.47 it would be considered a moderate correlation according to Bahna (2009). However, given that the value lies very close to the lower limit of moderate correlation values, this relationship is accepted. All other relationships between two independent variables are correlated with a value lower than 0.4 which is considered low as previously stated. Therefore, these relationships are all accepted.

5.3 Data Handling

The aim of this section is to provide readers with a clear understanding of how the data is handled from when it is collected from S&P Capital IQ to when the regression analysis is applied. The data handling is divided into handling of empty data, handling of outliers, and avoidance of the so-called dummy variable trap.

5.3.1 Handling Empty Data

Although multiple data fields were included for all transactions, many of these columns contained missing data, resulting in empty data points. Out of the initial 8,463 data points available, only the transactions with complete data in all the desired data fields were kept, i.e. the removal of rows with empty columns. Additionally, the 'Financials' sector was excluded as it did not contain any data points. After eliminating transactions with missing data or NaN values, 1,439 unique transactions remained in the dataset.

5.3.2 Outliers

The dataset contained transactions where the IEV was significantly higher than other observations of the same variable. In order to prepare the dataset for the regression analysis, these transactions were removed. Following Chatterjee and Hadi (1986) recommendations discussed in Section 3.4, outliers were removed only after using appropriate influence diagnostic tools. The data points to remove were chosen using the Z-score method. Values with a Z-score greater than +2 or less than -2 were considered outliers, in accordance with the common rule of thumb based on the properties of the normal distribution explained in Section 3.4.1. This resulted in a decrease from 1439 to 818 transactions left in the dataset. Further handling of outliers was done using the influential diagnostic tool Cook's distance. In accordance with the general guideline presented in Section 3.4.2, values greater than the threshold $\frac{4}{n}$, where n is the number of observations, were considered influential and removed. After removing these outliers, 785 transactions remained in the final version of the dataset. These outliers were excluded from the regression analysis to prevent them from disproportionately influencing the results.

5.3.3 Dummy Variable Trap

To avoid falling in to the dummy variable trap explained in Section 3.6.5, one of the binary sector variables was omitted. The excluded sector was communication services, transactions that occurred within this sector therefore acted as reference variables. Communication services was omitted on the basis of the alphabetical order of sectors, as to not skew the dataset.

5.4 Descriptive Data

The initial dataset consisted of 8463 transactions that ranged from the year 2005 to 2025. After being cleaned and prepared, as seen in Table 5.1 the final dataset consisted of 785 unique transactions.

Table 5.1: Descriptive Statistics of Key Financial Variables.

Variable	**Unit**	**Obs**	**Mean**	**Std**	**Min**	**25%**	**50%**	**Max**
IEV	M USD	785	921	453	300	507	842	1871
Market Cap	M USD	785	488	327	1	222	403	2036
Net Debt	M USD	785	275	235	1	106	230	1434
Total Assets	M USD	785	772	511	59	401	667	3923
EBITDA–EBIT	M USD	785	45	53	0	17	32	687
Revenue–EBITDA	M USD	785	672	855	-81	188	383	6188
Net Income Margin	%	785	-13	76	-929	-7	1	192
CFNAI	–	229	-0.07					

The IEVs ranged from $300 million on the low end to $1871 million. The mean of IEVs was $921 million, with a standard deviation of $453 million. As the quantiles suggest, there is a disproportionate amount of observation where the IEVs are in the lower range of the observed interval. The market capitalization ranges from $1 million to $2036 million, with a mean of $488 million and standard deviation of $327 million. In line with the distribution of the EV, the majority of market capitalizations fell within the lower bounds of its range. Similarly to the market capitalization, the lowest observed value of net debt was $1 million, but the highest observed value was $1434 million. Both the mean and the standard deviation of net debt were significantly lower than those of the variables mentioned above. As with IEV and market capitalization, the majority of net debts fell within the lower bounds.

In contrast to the variables mentioned above, the total assets exerted a far higher maximum value of $3923 million and a minimum value of $59 million. Despite the larger value span of total assets, the mean value of $772 million and the standard deviation of $511 million are in line with the previously mentioned variables. EBITDA-EBIT had a mean value of $45 million, standard deviation of $53 million. The minimum value within the dataset was $0 million, whereas the maximum value was $687 million. The maximum value was significantly higher than the 75-quantile at only $56 million. The mean and standard deviation of Revenue-EBITDA, are in line with above mentioned variables with $672 and $855 million respectively. In contrast the minimum value for Revenue-EBITDA was negative, yet the maximum value was large at $6188 million.

The net income margins within the final dataset had a rather large spread, with the smallest value of -929% and the largest at 192%. The mean value was -13%, with a standard deviation of 76%. Most net income margins where however clustered around 0%, with a 25-quantile at -7% and a 75-quantile at 5%. The control variable CFNAI-index, had a mean value of -0.07 over the studied period.

Table 5.2: Sector-wise Observations.

Sector	Number of observations
Consumer Discretionary	138
Consumer Staples	35
Energy	95
Health Care	114
Industrials	129
Information Technology	81
Materials	52
Real Estate	55
Utilities	12
Reference sector	74

In Table 5.2, the composition of the sections for the studied transactions is presented. Consumer discretionary, health care, and industrials account for approximately half the transactions. Whereas transactions within the sectors consumer staples and utilities compose a rather small part, with 35 and 12 observations respectively. The remaining four sectors were observed between 50 and 100 times, representing the remaining observations. The 74 observations labeled as reference sector, represent the transactions classified as communication services which, as explained in Section 5.3.3, were omitted.

5.5 Data Analysis

To improve the fit of the model and reduce heteroscedasticity, the dependent variable (IEV) was logarithmically transformed, using the natural logarithm. The data was then split into training and testing sets and a multiple linear regression was run using the OLS method. Model assumptions were carefully tested, including checks for linearity (via scatter plots), multicollinearity (using VIF), and homoscedasticity. Finally, the residuals were analyzed for normality using histograms and QQ-plots. The model's performance was evaluated by comparing predicted and actual values, and by calculating metrics such as MAE, MAPE, and R^2. When evaluating the model's performance, variables that turned out to be insignificant to the model's prediction were removed. The significance of variables was determined using the p-value as presented in Section 3.5.1 meaning that variables with a p-value higher than the chosen significance level of 0.05 were removed. Both the variables included to predict IEV and those included as control variables were systematically removed throughout the evolution of the model. This approach left only statistically significant variables in the final model.

6

Results and Analysis

In order to give nuanced and insightful results for analysis, a total of four different models were constructed. The first one included all the independent variables used in this thesis. Each model is presented with three tables, where the first table shows the slope coefficient, the standard error for each coefficient, and the p-value for each variable. The second table includes test statistics such as R^2, MAE, MAPE and AIC to facilitate performance comparison between the models. The third table displays the models performance on the B-P test, in order to determine whether the residuals are heteroscedastic or not. Moreover, the VIF score for each variable is also presented for all models except the first one, where the VIF score is presented in the Appendix due to the extensive number of variables used. In addition, the normality of the residuals is discussed for each model.

6.1 First Model

Table 6.1: Regression Results Model 1

Independent Variable	**Slope Coefficient**	**Std. Error**	**P-value**
Constant	5.715	0.027	0.000
Market Cap	0.001	2.56×10^{-5}	0.000
Net Debt	0.001	4.45×10^{-5}	0.000
Revenue–EBITDA	-1.90×10^{-5}	1.05×10^{-5}	0.071
Total Assets	0.0001	2.57×10^{-5}	0.000
Net Income Margin	-0.012	0.010	0.211
EBITDA–EBIT	-2.60×10^{-5}	0.000	0.871
CFNAI	-0.034	0.014	0.018
Consumer Discretionary	-0.0008	0.028	0.977
Consumer Staples	0.107	0.041	0.010
Energy	-0.040	0.031	0.204
Health Care	0.055	0.030	0.069
Industrials	0.039	0.029	0.186
Information Technology	0.028	0.033	0.405
Materials	0.035	0.036	0.335
Real Estate	-0.001	0.035	0.970
Utilities	-0.096	0.071	0.174

Note: For test statistics see Table 6.2

As seen in Table 6.1, there are many insignificant variables. Due to this, it seems plausible that the model can be simplified by reducing the number of variables and still maintain approximately the same explanatory level in accordance with Section 3.7.3.

Table 6.2: Test Statistics Model 1

Metric	**Value**
R^2	0.71
MAE	163.71
MAPE	18.52%
AIC	-373.5

Table 6.3: B-P Test Results Model 1

Statistic	**Value**
Lagrange Multiplier	24.574
P-value	0.078
F-value	1.555
F P-value	0.076

In Table 6.2 it is shown that the model has an R^2 value of 0.71 together with an MAE of 163.7 and an MAPE of 18.5%. Furthermore, the AIC was -373.5. As seen when visually inspecting Appendix J, the model is more accurate for companies with a lower EV than for companies with a higher EV.

Table 6.3 concluded that the data were homoscedastic. Also, as seen in Appendix A, all variables passed the multicollinearity test since their VIF scores are below

10. Moreover, as seen in Appendices B and F, the residuals are seemingly normally distributed, meaning that they fulfill another one of the key assumptions for the analysis.

6.2 Second Model

Table 6.4: Regression Results Model 2

Independent Variable	Slope Coefficient	Std. Error	P-value
const	5.695	0.027	0.000
Market Cap	0.001	2.480×10^{-5}	0.000
Net Debt	0.001	4.320×10^{-5}	0.000
Total Assets	0.0001	2.090×10^{-5}	0.000
CFNAI	-0.039	0.014	0.007
Consumer Discretionary	-0.006	0.028	0.826
Consumer Staples	0.103	0.045	0.022
Energy	-0.040	0.031	0.194
Health Care	0.068	0.030	0.023
Industrials	0.037	0.029	0.195
Information Technology	0.031	0.032	0.338
Materials	0.017	0.036	0.639
Real Estate	-0.011	0.035	0.748
Utilities	-0.021	0.076	0.782

Note: For test statistics see Table 6.5

As stated previously, the model should have approximately the same explanatory power after removing the insignificant variables in the first model. Therefore, this was performed for the second model, and the insignificant variables were omitted. However, as shown in Table 6.4, the model still consists of several insignificant variables.

Table 6.5: Test Statistics Model 2

Metric	Value
R^2	0.47
MAE	172.67
MAPE	18.21%
AIC	-379.4

Table 6.6: B-P Test Results Model 2

Statistic	Value
Lagrange Multiplier	20.714
P-value	0.079
F-value	1.611
F P-value	0.077

Table 6.5 shows that the model had a R^2 value of 0.47, an MAE of 172.7, an MAPE of 18.2%, and an AIC score of -379.4. Compared to the first model, all metrics except R^2 performed better in the second model. In Appendix K, it is easily seen

that the accuracy of the model is higher for companies with a lower EV, although there exist some outliers for the lower-valued companies.

Table 6.6 shows that the model was homoscedastic. Another important note is that all variables were well below the highest acceptable VIF score as shown in Table 6.7, which means that there was no multicollinearity.

Table 6.7: VIF Scores Model 2

Independent variable	VIF
const	14.27
Market Cap	1.33
Net Debt	2.06
Total Assets	2.23
CFNAI	1.03
Consumer Discretionary	2.42
Consumer Staples	1.31
Energy	1.93
Health Care	2.21
Industrials	2.28
Information Technology	1.90
Materials	1.60
Real Estate	1.63
Utilities	1.10

However, as stated in Section 3.7.3, some control variables might be worth omitting due to both statistical insignificance and a small sample. Therefore, this was done in an effort to simplify the model without losing explanatory power. Regarding the normal distribution, it is obvious from visual inspection of Appendices C and G that the residuals follow the normal distribution reasonably well and therefore meet the assumption of normality.

6.3 Third Model

Table 6.8 shows that for the third model, not all variables that had previously been significant were significant. In fact, the two control variables CFNAI and the consumer staples sector turned out to be insignificant.

Table 6.8: Regression Results Model 3

Independent Variable	Slope Coefficient	Std. Error	P-value
const	5.715	0.017	0.000
Market Cap	0.001	2.60×10^{-5}	0.000
Net Debt	0.001	4.33×10^{-5}	0.000
Total Assets	0.0001	2.28×10^{-5}	0.000
CFNAI	-0.029	0.015	0.054
Consumer Staples	0.065	0.036	0.069
Health Care	0.050	0.022	0.023

Note: For test statistics, see Table 6.9

Model 3 performed on a level similar to the previous models in the statistical tests. As shown in Table 6.9, the model had a R^2 of 0.7, an MAE of 165.7, an MPAE of 18.3% and an AIC score of -376.7. Furthermore, Appendix L indicates that the model performs significantly better for companies with a lower EV.

Table 6.9: Test Statistics Model 3

Metric	Value
R^2	0.70
MAE	165.7
MAPE	18.3%
AIC	-376.7

Table 6.10: B-P Test Results Model 3

Statistic	Value
Lagrange Multiplier	10.529
P-value	0.104
F-value	1.765
F P-value	0.104

As observed in Table 6.10, the third model was homoscedastic as well. Furthermore, all included variables passed the multicollinearity test, as they were well below 10, which can be seen in Table 6.11. Furthermore, the residuals appear to be normally distributed when examining Appendices D and H visually. Due to the fact that the model still consists of a couple of insignificant variables, an additional model was tested, where the remaining insignificant variables were omitted.

Table 6.11: VIF Scores Model 3

Independent variable	VIF
const	5.43
Market Cap	1.40
Net Debt	1.94
Total Assets	2.24
CFNAI	1.01
Consumer Staples	1.01
Health Care	1.07

6.4 Fourth Model

In the fourth and final model it is clear from Table 6.12 that all included variables were statistically significant. As for all previous models, the slope coefficient and standard error can be seen in Table 6.12 together with the p-value for each variable.

Table 6.12: Regression Results Model 4

Independent Variable	Slope Coefficient	Std. Error	P-value
const	5.718	0.016	0.000
Market Cap	0.001	2.59×10^{-5}	0.000
Net Debt	0.001	4.30×10^{-5}	0.000
Total Assets	0.0001	2.26×10^{-5}	0.000
Health Care	0.043	0.022	0.047

Note: For test statistics see Table 6.13

As shown in Table 6.13, the final model got a R^2 of 0.7, an MAE of 160.0, an MAPE of 17.8%, and an AIC score of -391.8. Compared to the previous models, the different statistics have remained fairly the same for all four. In Appendix M, it is implied that the model performs substantially better for companies with lower EVs, as all the previous models also did.

Table 6.13: Test Statistics Model 4

Metric	Value
R^2	0.70
MAE	160.0
MAPE	17.8%
AIC	-391.8

Table 6.14: B-P Test Results Model 4

Statistic	Value
Lagrange Multiplier	4.429
P-value	0.351
F-value	1.106
F P-value	0.353

Regarding the B-P test, shown in Table 6.14, it is obvious that the model once again turned out to be homoscedastic. Furthermore, the VIF score for each variable shown in Table 6.15 indicates the absence of multicollinearity. It is also clear when studying Appendices E and I that the residuals of the model fit the normal distribution ostensibly well, leading to the assumption of normally distributed residuals being satisfied.

Table 6.15: VIF Scores Model 4

Independent variable	VIF
const	5.29
Market Cap	1.40
Net Debt	1.95
Total Assets	2.26
Health Care	1.06

The final model presented in Table 6.12 generates an equation for calculating IEV as shown in Equation 6.1 where X_1 is market capitalization, X_2 is net debt, X_3 is total assets, and X_4 is the health care sector.

$$\text{IEV} = \exp(5.718 + 0.001 \times X_1 + 0.001 \times X_2 + 0.0001 \times X_3 + 0.043 \times X_4) \quad (6.1)$$

7

Discussion

7.1 Model Evaluation

When evaluating which model is the most valid out of the four presented in the results, it is important to consider both the test statistics and the theoretical importance of certain variables. Therefore, since some control variables turned out to be statistically insignificant, these variables need to be thoroughly examined to decide whether they should be part of the final model or not.

7.1.1 Control Variables Significance

Given that so few control variables turned out to be statistically significant, one might question if they were not needed or if they did not control the desired aspect. As previously stated, the original model consisted of control variables for the size of the companies, the current state of the market, and for the different sectors in which the companies act. When judging solely by the statistical tests, most sectors turned out to be insignificant, as well as the CFNAI variable, which was used to control for the current state of the market. This can be somewhat surprising considering that the valuation multiples often differ between sectors as stated in Section 1.2 and that would imply that sector is significant in determining the EV. One potential reason for most sectors being statistically insignificant may be that the sample from some sectors was too small to make any overall difference. However, as seen in Table 6.8, the consumer staples sector is close to being statistically significant together with the health care sector, which might indicate greater differences in valuation for these specific sectors.

Regarding CFNAI, it had a p-value that indicated that it was nearly statistically significant in the fourth model, while it was statistically significant in all other models. As CFNAI is used to control the current state of the US market and appears to have an effect on the EV of a company, as stated in Section 2.9, it has a great theoretical importance. Considering the great theoretical importance of CFNAI as a control variable and the fact that it was very close to being statistically significant, it is, in accordance with the reasoning in Section 3.7.3, plausible that the most valid model should include CFNAI as a variable. As mentioned above, the control variable for the size of the company was the market capitalization, and this turned out to be a statistically significant control variable. The fact that the market capitalization was so effective also seems quite intuitive given how EV is calculated in Equation 2.1.

7.1.2 Examining the Model

When determining which model performs the best of the four, several aspects must be taken into account. Firstly, there are several different measurements regarding model performance that need to be evaluated. With respect to R^2, the second model performs the worst, and the other three models perform very similarly to each other. Given the definition of MAE in Section 3.8.1, MAE should be as low as possible, as this indicates a smaller error. Considering that, model four seems to perform the best in absolute terms, although there is a small difference between the models. However, one must also recognize that since the data were split into a train and a test set randomly by default, MAE might be somewhat deceiving, as it will be lower if the split of the dataset merely gives a higher proportion of lower-valued companies. Therefore, MAPE is also of interest, as it gives error in terms of percentage instead of absolute values. Examining the MAPE for each model shows that model four once again performed the best, although the differences were not very large between the different models. Also, when inspecting the AIC model four performed the best, whereas the other models performed very similarly to each other.

Considering all these test statistics and since all models seemed to fulfill the general assumptions for an OLS regression analysis, it is tempting to conclude that model four was the best model. However, there are other aspects to consider before determining whether it is the best model. Such aspects mainly include the validity of dropping control variables if they are insignificant. As stated in Section 3.7.3 the omission of insignificant control variables is a twisted subject, there are some reasons that strengthen the case of omitting them and some reasons to include them. When taking into account how close the third model was to being statistically significant, the relatively even test results, and the unsolved argument regarding keeping insignificant control variables, one might argue that this was the most valid one of the four. To conclude the model evaluation, both the third and the fourth model performed very well, but considering the arguments regarding keeping nearly significant control variables with a great theoretical importance, the third model was deemed the best.

7.1.3 Model and Data Limitations

Despite the extensive dataset, several important data omissions limit the explanatory power of the model. First, the dataset lacks historical growth indicators, such as EBITDA CAGR, which prevents our model's ability to take into account long-term performance trends that often influence valuation. In addition, forward-looking estimates, such as forecasted financials or analyst projections, are absent limiting our model's insights into market expectations and anticipated future growth.

A further limitation is the omission of strategic and qualitative factors. Important deal factors, such as expected synergies and integration challenges are not captured, despite their relevance in M&A valuation. Moreover, the dataset is restricted to publicly traded targets with disclosed financials, introducing potential selection bias. This may result in survivorship and reporting bias due to excluding, for instance, private transactions that could provide meaningful contrasts and insights.

From a methodological standpoint, the model relies on OLS regression, which assumes linear relationships between independent variables and EV. This may be an oversimplification, particularly if the true relationships are non-linear or involve other external effects.

7.2 Influential Variables

The results indicate that net debt and total assets are important in explaining the independent variable, IEV. This is due to the fact that both were found to be statistically significant. This supports earlier research and the economic theory discussed in Section 2, which suggests that net debt and total assets affect a company's value. Net Debt influences EV directly as its part of the equation of EV and indirectly though representing the company's risk profile. Total assets are also important because larger companies often benefit from economies of scale and more stable performance, which usually leads to a higher EV.

In contrast, some variables that were expected to be important turned out to be statistically insignificant. Net income margin, which measures profitability, was not found to be significant in explaining IEV, although economical theory suggested it is. The same applies to both the proxy variable Revenue-EBITDA which was used to represent operating costs and COGS, and the proxy variable EBITDA-EBIT, which was used to represent amortization and depreciation. These results suggest that measurements related to expenses and profitability do not seem to be important for estimating IEV in M&A transactions.

When discussing why these variables turned out to be insignificant, it is important to consider that the studies presented in the economic theory do not specifically examine how these variables affect IEV in M&A transactions. As mentioned in the introduction, the factors that influence IEV in an M&A context have not been fully explored in earlier research. A lot of the existing studies on EV are more focused on general valuation or stock market settings, where things like profitability and cost structure might have more of an effect. But in M&A deals, other factors can matter more. Because of this, variables like net income margin or the proxies used for COGS, operational costs, amortization, and depreciation might not show up as significant in the results. Their effect could be smaller or even overshadowed by other deal-specific aspects that are more important when companies are being bought or merged.

Additionally, an interesting observation from the results is that the insignificant variables, namely net income margin, Revenue-EBITDA, and EBITDA-EBIT, are all derived from the income statement, while the significant variables, such as net debt and total assets, are connected to the balance sheet. This suggests that, in determining IEV in an M&A transaction, variables from the balance sheet may carry more weight than those from the income statement. This leads to a natural question regarding the relative importance of income statement data compared to balance sheet data in such transactions. As Sorrentino et al. (2024) explains, the balance sheet reflects the entire structure of a company's assets and liabilities, not just its profitability in a certain year, as compared to the income statement. Thus, the balance sheet is crucial for understanding a company's financial strength during M&A. Moreover, Sorrentino et al. (2024) highlights that items related to total assets, such as how intangible assets like goodwill are recorded, can significantly influence perceived company value and provide insights into its future earnings potential.

One reason why balance sheet variables may have appeared more influential in this thesis is that the data was taken at the announcement date rather than averaged over several years. The balance sheet at a specific point in time offers a clearer picture of the company's overall financial health, while the income statement reflects performance over the most recent year only. This difference is important, as noted by Kpentey (2015), who points out that income can vary significantly from year to year. A single years result is often not enough to determine sustainable earnings because it can be affected by many short-term factors. Supporting this view, Sloan (1996) explains that companies sometimes report unusually high or low earnings and costs due to special events. For example, one-time expenses such as restructuring costs or one-time gains from asset sales may distort the company's actual earning ability.

Therefore, in the context of M&A, especially when using values recorded on the announcement date, the balance sheet may be a more reliable indicator of a company's true value than the income statement. It reflects the structure of a company's assets and liabilities more clearly, without being affected by short-term fluctuations. This could explain why the balance sheet variables were found to be more significant in this research. However, if sustainable revenues or costs had been calculated as averages over several years leading up to the acquisition, income statement variables might have gained more relevance. In this way, the timing and scope of the financial data play a key role in determining which metrics are most useful for evaluating a company in M&A transactions.

7.3 Practical Use of the Results

The results offer a potential complement to existing and commonly used valuation methods such as DCF and CCA. Although the results imply that the model does not determine the IEV completely accurately, it gives a good benchmark for the IEV. As seen in Appendices J, K, L, and M, the model performed worse for higher-valued companies, and there were some outliers regarding the accuracy of the model. Considering this, the model could potentially work as an even better indicator of IEV than expected from its test statistics for companies that should be valued at a lower level. Given that the IEV was distributed mainly in the lower range, as stated in Section 5.4, the higher accuracy of the model is expected for companies with lower values. However, it is hard to know beforehand whether a company should be higher- or lower-valued. Overall, the model can give a generally accurate benchmark of the IEV in a much faster and easier way than the most widely used valuation methods as of now and could therefore serve as a good complement to these.

For the aspect of the results regarding what factors that affect the IEV, it is obvious that net debt, market capitalization, total assets, and whether the company acted in the health care sector affect the value. In some sense, the CFNAI, as well as whether the company acted in the consumer staples sector, also affects the IEV considering that they were so close to being statistically significant. These factors do not have the same clear practical potential as the model and the usage of it, but it is inaccurate to state that it has no practical usage as it gives obvious factors that affect the value of a company.

7.4 Continuous Research

As stated in Section 1.5 and 5.1, this thesis is conducted on small-cap companies based in the U.S. Therefore, future research on which factors influence IEV in M&A transactions could benefit from examining firms of different sizes or in other geographical contexts. For instance, studies focusing on specific regions such as the Nordics, Asia, or Africa could provide valuable insights, as the factors influencing IEV may vary across markets due to differences in regulatory environments, economic structures, or market maturity.

Given the limited existing research on the topic, as discussed in Section 1.1, there is uncertainty regarding which factors influence IEV and also how these factors differ between regions or company sizes. This lack of clarity presents a valuable opportunity for future studies to build a deeper understanding of how external conditions such as regulation, market maturity, or regional characteristics affect IEV in M&A transactions.

Another area to consider for future research is the inclusion of additional factors that may influence IEV, such as specific components of the balance sheet or qualitative deal factors. For example, as mentioned in Section 7.2, total assets include goodwill, which could impact the final IEV in M&A transactions. In addition, strategic elements such as expected synergies or the complexity of integrating the acquired organization into the existing structure may also play an important role. These qualitative aspects can be translated into proxy variables, allowing their effects on IEV to be tested and evaluated alongside traditional financial metrics. Incorporating such variables into regression models could bridge the gap between quantitative data and strategic insights, resulting in a more complete understanding of value drivers in M&A deals.

8

Conclusion

In accordance with its purpose, this thesis has accentuated the factors affecting the valuation of companies in M&A transactions. This has been achieved through the development, refinement, and testing of an OLS model. Moreover, the following research questions have been used to fulfill the studies purpose.

- Which financial variables significantly influence EV in M&A transactions?
- How can these variables be integrated into a regression model to predict EV?

The thesis concludes that net debt and total assets are the two financial variables that exert the greatest influence on IEV within M&A transactions. The regression analysis also shows that net income margin, EBITDA-EBIT and revenue-EBIT are insignificant and do not considerably contribute to explaining the IEV in M&A-transactions. These findings indicate that items in a company's balance sheet are strong predictors of the company's IEV whereas items on its income statement are not, at least in their form in the conducted study. Furthermore, the control variable market capitalization influenced the IEV significantly. Exhibiting the same slope coefficient as net debt in all regression models.

The regression analysis conducted in this thesis, yielded a formula for calculating the IEV of a small-cap U.S. M&A transaction. The formula uses the target company's market capitalization, net debt, total assets, and sector as key inputs. Usage of a formula offers a fast and easy method for benchmarking a company's valuation in the context of M&A transactions. Through the use of readily available financial metrics such as market capitalization, net debt and total assets, the regression model can supplement more conventional valuation methods and give valuable insights to actors within M&A.

Future research should examine the influence of balance sheet components, such as goodwill, on EV, testing whether variations in goodwill levels correlate with valuation outcomes across M&A transactions. This could be done using a quantitative approach similar to the one conducted in this thesis. Studies should also look beyond the U.S. exploring other regions and firms of different sizes to assess how regulatory, economic, or market maturity differences affect IEV. Another aspect that should be further explored is the inclusion of rarely quantified qualitative deal attributes, potentially capturing the impact of variables such as expected synergies on valuation in M&A transactions.

Bibliography

Akhtar, W. M., Khan, F. A., Shahid, A., & Ahmad, J. (2016). Effects of debt on value of a firm. *Journal of Accounting & Marketing*, *5*(4), 202. https://doi.org/10.4172/2168-9601.1000202

Bahna, S. L. (Ed.). (2009). *Statistics for clinicians* (Vol. 103) [Supplement to October 2009 Issue]. Annals of Allergy, Asthma & Immunology, Supplement 1.

Barnes, B., Harp, N., & Oler, D. (2014). *Evaluating the sdc mergers and acquisitions database* (tech. rep.). SSRN Electronic Journal. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2201743#:~:text=SDC%20data%20across%20time,deal%20with%20important%20subtleties%20in

Barnett, V., & Lewis, T. (1994). *Outliers in statistical data* (3rd ed.). Wiley.

Blackham, R. (2023). Valuation models & discount rates: An in-depth analysis of the dcf-model & wacc [CBS HD(R) Final Project, Copenhagen Business School].

Brave, S. A. (2020). A closer look at the chicago fed's activity indexes. In C. J. Hueng (Ed.), *Alternative economic indicators* (pp. 25–40). W.E. Upjohn Institute. https://books.google.se/books?id=qNT7DwAAQBAJ&pg=PA25

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, *47*(5), 1287–1294. https://doi.org/10.2307/1911963

Brockwell, P. J., & Davis, R. A. (2016). *Introduction to time series and forecasting* (3rd ed.). Springer. https://doi.org/10.1007/978-3-319-29854-2

Cavanaugh, J. E., & Neath, A. A. (2019). The akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *Wiley Interdisciplinary Reviews: Computational Statistics*, *11*(3), e1460. https://doi.org/10.1002/wics.1460

Chatterjee, S., & Hadi, A. S. (1986). Influential observations, high leverage points, and outliers in linear regression. *Statistical Science*, *1*(3), 379–416. https://doi.org/10.1214/ss/1177013622

Chen, J. (2020). *Comparable company analysis (cca): How is it used in investing?* [Accessed: 2025-04-21]. Investopedia. https://www.investopedia.com/terms/c/comparable-company-analysis-cca.asp

Cook, R. D. (1977). Detection of influential observations in linear regression. *Technometrics*, *19*(1), 15–18. https://doi.org/10.1080/00401706.1977.10489493

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications, Inc.

Damodaran, A. (2005). Valuation approaches and metrics: A survey of the theory and evidence. *Foundations and Trends in Finance*, *1*(8), 693–784. https://doi.org/10.1561/0500000013

Damodaran, A. (2012). *Investment valuation: Tools and techniques for determining the value of any asset* (3rd ed.). John Wiley & Sons.

Dang, H. N., Vu, V. T. T., Ngo, X. T., & Hoang, H. T. V. (2019). Study the impact of growth, firm size, capital structure, and profitability on enterprise value: Evidence of enterprises in vietnam. *Journal of Corporate Accounting & Finance*, *30*(1), 144–152. https://doi.org/10.1002/jcaf.22371

Das, B., & Resnick, S. I. (2008). Qq plots, random sets and data from a heavy tailed distribution. *Stochastic Models*, *24*(1), 103–132. https://doi.org/10.1080/15326340701828308

Denscombe, M. (2010). *The good research guide for small-scale social research projects* (4th ed.). McGraw-Hill Education.

Elmberger, H., & Makdisi-Somi, M. (2016). *Using regression analysis to determine the enterprise value of a company: A regression analysis on the enterprise value of companies within the industry manufacturing of chemicals and chemical products* [Degree Project in Applied Mathematics and Industrial Economics]. KTH Royal Institute of Technology. https://kth.diva-portal.org/smash/get/diva2:942663/FULLTEXT01.pdf

Fernandez, P. (2007). *Company valuation methods: The most common errors in valuations* (tech. rep. No. 449). IESE Business School. https://media.iese.edu/research/pdfs/DI-0449-E.pdf

Fernando, J. (2024). *Enterprise value (ev) formula and what it means* [Accessed: 2025-04-21]. Investopedia. https://www.investopedia.com/terms/e/enterprisevalue.asp

Hayashi, F. (2000). *Econometrics*. Princeton University Press.

Hermuningsih, S. (2013). Profitability, growth opportunity, capital structure and the firm value. *Bulletin of Monetary Economics and Banking*, *16*(2), 115–130. https://doi.org/10.21098/bemp.v16i2.440

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in r*. Springer. https://doi.org/10.1007/978-1-4614-7138-7

Juárez, F. (2018). The growth of companies as a function of total assets. *WSEAS Transactions on Business and Economics*, *15*, 301–310.

Kaiser, M. J., & Snyder, B. F. (2013). Offshore driller valuation models. In *Lecture notes in energy* (pp. 139–148, Vol. 8). Springer-Verlag London. https://doi.org/10.1007/978-1-4471-5152-4_8

Kaplan, S. N., & Ruback, R. S. (1995). The valuation of cash flow forecasts: An empirical analysis. *The Journal of Finance*, *50*(4), 1059–1093. https://doi.org/10.1111/j.1540-6261.1995.tb04050.x

Kara, H. (2018). *Research ethics in the real world*. Policy Press.

Kenton, W. (2024). *Sarbanes-oxley (sox) act of 2002 definition* [Accessed: 2025-05-01]. Investopedia. https://www.investopedia.com/terms/s/sarbanesoxleyact.asp

Kpentey, B. (2015). Income approach: Capitalization methods [Chapter contributed by Sync Consult Limited, Ghana and University of Ghana]. In *Wiley guide to fair value under ifrs: International financial reporting standards* (pp. 57–64). John Wiley & Sons. https://doi.org/10.1002/9781119204008.ch4

Lang, A. (2014). *Introduction to regression analysis* [Accessed: 2025-04-24]. https://people.kth.se/~lang/regression_analysis.pdf

Lino, G. (2024). *A comprehensive guide to valuation using multiples: Understanding the method and its applications*. https://www.girolino.com/a-comprehensive-guide-to-valuation-using-multiples-understanding-the-method-and-its-applications/

MacDiarmid, J., Tholana, T., & Musingwini, C. (2018). Analysis of key value drivers for major mining companies for the period 2006–2015. *Resources Policy*, *56*, 16–30. https://doi.org/10.1016/j.resourpol.2017.09.008

Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive statistics and normality tests for statistical data. *Annals of Cardiac Anaesthesia*, *22*(1), 67–72. https://doi.org/10.4103/aca.ACA_157_18

Ouimet, P. P. (2013). What motivates minority acquisitions? the trade-offs between a partial equity stake and complete integration. *Review of Financial Studies*, *26*(4), 1021–1047. https://doi.org/10.1093/rfs/hhs147

Sarstedt, M., & Mooi, E. (2018). Regression analysis. In *A concise guide to market research* (pp. 209–256). Springer. https://doi.org/10.1007/978-3-662-56707-4_7

Schoenmaker, D., & Schramade, W. (2023). *Corporate finance for long-term value*. Springer. https://doi.org/10.1007/978-3-031-35009-2

Schueler, A. (2020). Valuation with multiples: A conceptual analysis. *Journal of Business Valuation and Economic Loss Analysis*. https://doi.org/10.1515/jbvela-2019-0020

Shen, Z., & Fan, Z. (2017). An empirical study of period expense on operational efficiency of listed manufacturing enterprises in china. *Boletin Tecnico/Technical Bulletin*, *55*(15), 554–563.

Sloan, R. G. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review*, *71*(3), 289–315. http://www.jstor.org/stable/248290

Sorrentino, G., Situm, M., Serpeninova, Y., Tumpach, M., & Juhaszova, Z. (2024). Supporting management decisions for m&a transactions based on the strategic allocation of intangible assets [Gold Open Access]. *Problems and Perspectives in Management*, *22*(2), 539–554. https://doi.org/10.21511/ppm.22(2).2024.42

S&P Global Market Intelligence. (2021). *Calculation of implied enterprise value* [Accessed: 2025-05-01]. https://spglobal.my.site.com/s/article/Calculation-of-Implied-Enterprise-Value-on-the-Market-Intelligence-platform-MI

Suliyanto, Sediono, & Rosyada, A. (2019). Profitability level modelling of manufacturing companies based on binary logistic regression with random effect on panel data. *International Journal of Innovation, Creativity and Change*, *5*(3), 199–216.

Sureiman, O., & Mangera, C. M. (2020). F-test of overall significance in regression analysis simplified. *Journal of the Practice of Cardiovascular Sciences*, *6*(2), 116–122. https://doi.org/10.4103/jpcs.jpcs_18_20

Thompson, C. G., Kim, R. S., Aloe, A. M., & Becker, B. J. (2017). Extracting the variance inflation factor and other multicollinearity diagnostics from typical regression results. *Basic and Applied Social Psychology*, *39*(2), 81–90. https://doi.org/10.1080/01973533.2016.1277529

U.S. Securities and Exchange Commission. (2024). *Officers, directors and 10% shareholders* [Accessed April 2025]. U.S. Securities and Exchange Commission. https://www.sec.gov/smallbusiness/goingpublic/officers-directors-10percent-shareholders

Wallén, G. (1996). *Vetenskapsteori och forskningsmetodik* (2nd ed.). Studentlitteratur.

Watts, V. (2022). *Introduction to statistics: An excel-based approach* [Licensed under Creative Commons BY-NC-SA 4.0]. Fanshawe College Pressbooks. https://ecampusontario.pressbooks.pub/introstats/

Westlund, J., & Petrova, Y. (2018). Asymptotic collinearity in cce estimation of interactive effects models. *Economic Modelling*, *70*, 331–337. https://doi.org/https://doi.org/10.1016/j.econmod.2017.07.023

Wooldridge, J. M. (2013). *Introductory econometrics: A modern approach* (5th ed.). South-Western Cengage Learning.

Yin, R. K. (2003). *Case study research: Design and methods* (3rd ed.). SAGE Publications.

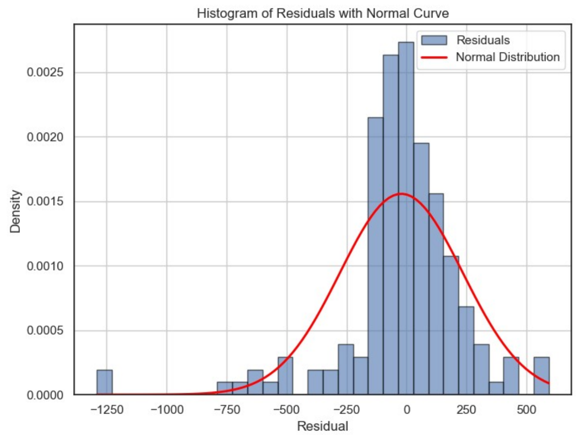
Zheng, X., Hu, M., Lai, L., & Li, J. (2023). Enterprise value evaluation of the byd company based on fcff model. In R. Gupta, F. Bartolucci, V. N. Katsikis, & S. Patnaik (Eds.), *Recent advancements in computational finance and business analytics* (pp. 529–544). Springer. https://doi.org/10.1007/978-3-031-38074-7_46

Appendix

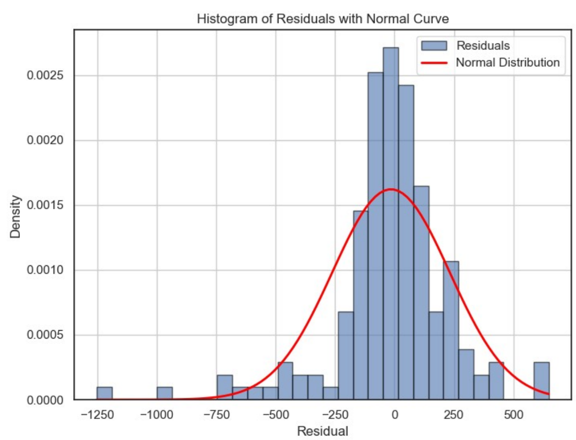
A: VIF Model 1

Independent variable	**VIF**
const	14.92
Market Cap	1.43
Net Debt	2.12
Revenue-EBITDA	1.60
Total Assets	3.12
Net Income Margin	1.11
EBITDA-EBIT	1.40
CFNAI	1.04
Consumer Discretionary	2.46
Consumer Staples	1.46
Energy	1.93
Health Care	2.23
Industrials	2.29
Information Technology	1.82
Materials	1.62
Real Estate	1.76
Utilities	1.11

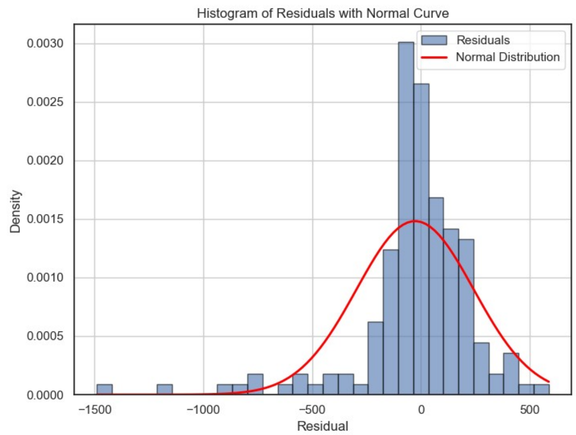
B: Histogram Model 1



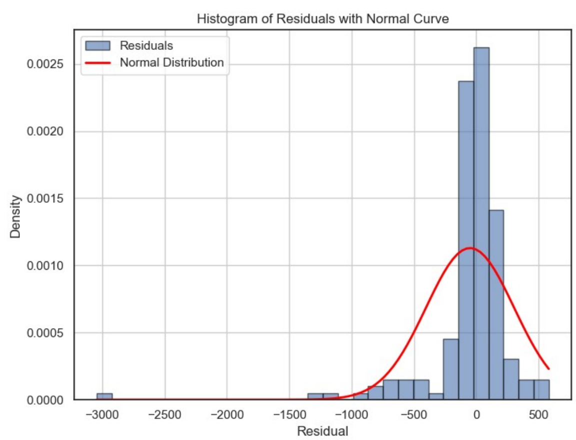
C: Histogram Model 2



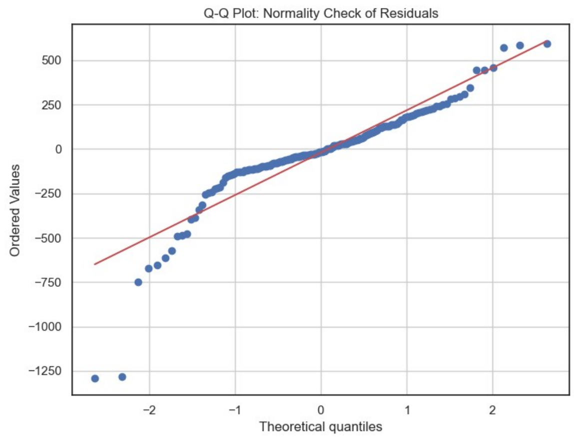
D: Histogram Model 3



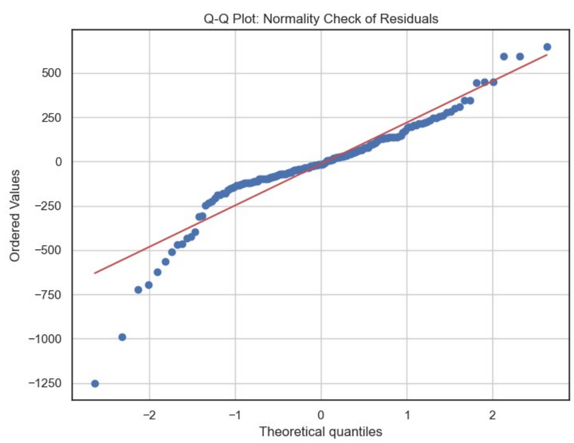
E: Histogram Model 4



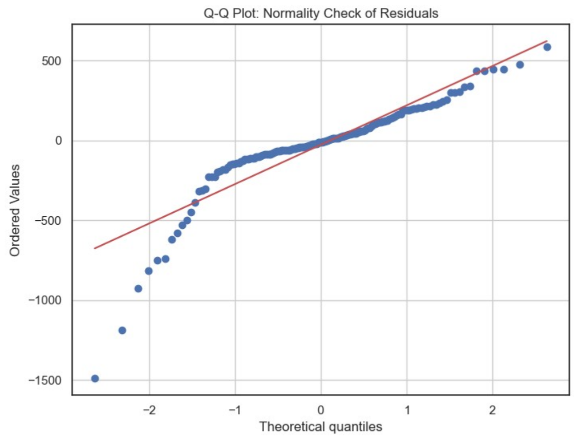
F: QQ-plot Model 1



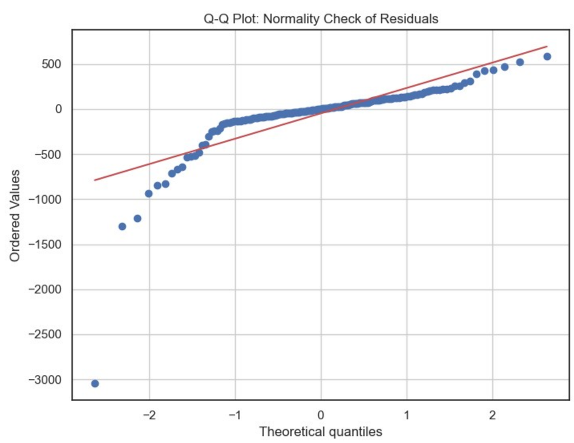
G: QQ-plot Model 2



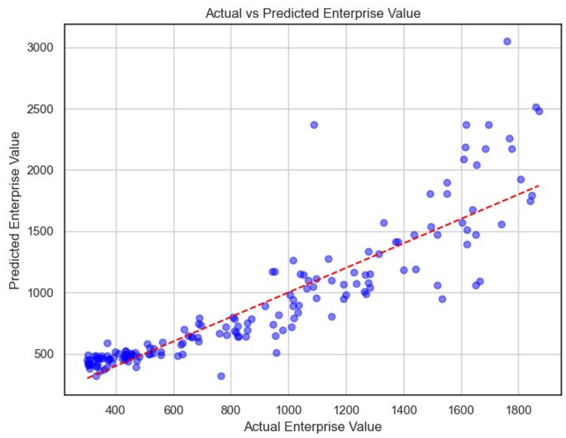
H: QQ-plot Model 3



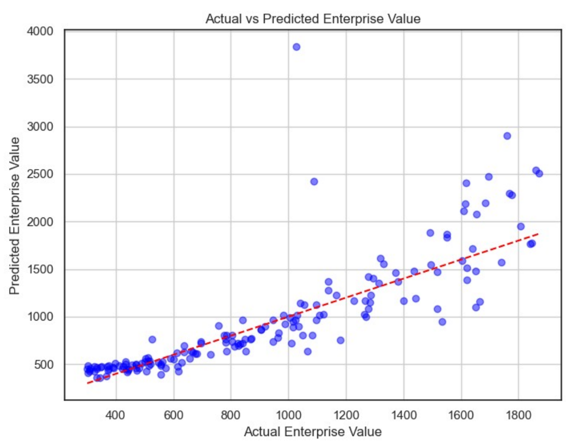
I: QQ-plot Model 4



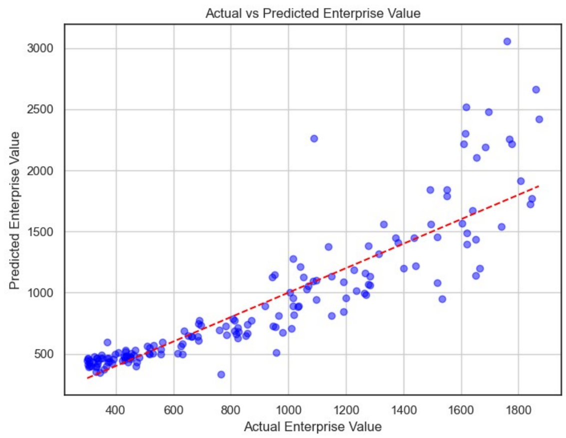
J: Actual vs Predicted EV model 1



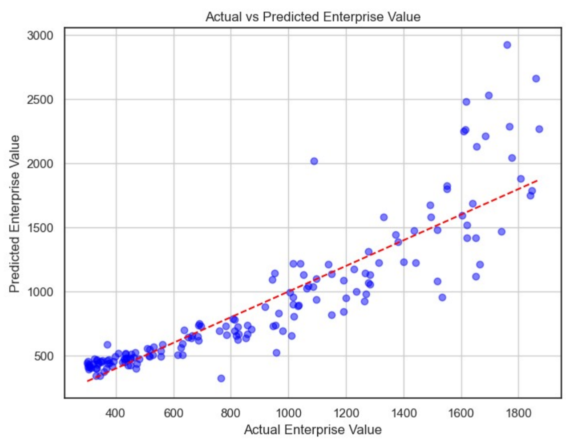
K: Actual vs Predicted EV model 2



L: Actual vs Predicted EV model 3



M: Actual vs Predicted EV model 4



DEPARTMENT OF TECHNOLOGY MANAGEMENT AND ECONOMICS
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

AVANCEZ
1829

CHALMERS