A Regression Analysis of the Parameters Influencing the Share Price During the First Day After an IPO

DANIEL JOHNSON & MATILDA CADDEO
Sammanfattning

Denna studie fokuserar på börsintroduktioner (IPO), processen i vilken ett företags aktier görs tillgänglig för allmän handel på börsen eller en annan marknadsplats. Trots omfattande globala osäkerheter under senaste åren, så har efterfrågan på börsnoteringar varit hög. Många av dessa IPOs har upplevt en positiv aktietillväxt under första dagen som ett publikt handlat bolag.


Resultatet av studien visar på stora brister i modellernas prestationsförmåga. Troligen grundar sig deras oförmåga i både aktiemarknadens komplexitet samt svårigheter i att kvantifiera och fanga relevanta faktorer som påverkar de initiala aktieprisfluktuationerna efter en IPO. Den reducerade modellen presterar dock något bättre och indikerar att variabler som NGM SME, First North, PreSubscribed och AmountNewlyIssued förklarar det mesta av variansen i den beroende variabeln.

Nyckelord: Regressionsanalys, Börsnotering, Aktier, Marknadsplats, Under Pricing, Investment Banking, Predicera
Abstract

This study focuses on initial public offerings (IPOs), which are the process of making a company’s shares available for public trading on a stock market. Despite global uncertainties in recent years, there has been a high demand for company listings in the market. Many IPOs have experienced a positive trend in share prices on the first day of trading as a publicly traded company.

The objective of this study is to develop a multiple linear regression model to analyze the impact of various parameters on the first day return of IPOs. The generated model will be evaluated to create a reduced model with an optimal subset of variables. The study will specifically focus on IPOs listed on some Nordic marketplaces during the period 2017-2022.

The results of the study suggest that the created models are not effective in capturing the variance of first day returns. The deficiency of the created models is likely due to both the complexity of the stock market and the difficulties of quantifying and capturing some of the factors impacting the initial performance of IPOs. However, the reduced model performs slightly better and indicates that variables such as NGM SME, First North, PreSubscribed, and AmountNewlyIssued explain most of the variance in the response variable.

Keywords: Regression Analysis, IPO, Stocks, Marketplace, Underpricing, Investment Banking, Prediction
## Contents

1 Introduction .......................................................... 1
   1.1 Background .................................................. 1
   1.2 Aim of the study ............................................. 1
   1.3 Research Question and Hypothesis ......................... 2
   1.4 Scope and Limitations ...................................... 2
   1.5 Previous Research .......................................... 3

2 Method ................................................................. 4
   2.1 Multiple Linear Regression ................................. 4
   2.2 Response Variable .......................................... 4
      2.2.1 Development in Stock Price During the First Day After
           an IPO .................................................. 4
   2.3 Regressor Variable ......................................... 5
      2.3.1 Date of IPO ............................................ 5
      2.3.2 Marketplace ........................................... 5
      2.3.3 Advisor ............................................... 6
      2.3.4 Offering .............................................. 6
      2.3.5 Amount Newly Issued ................................ 6
      2.3.6 Initial Share Price .................................. 6
      2.3.7 Market Capitalization ................................. 6
      2.3.8 Amount of Shares Pre Subscribed ...................... 7
      2.3.9 SPAC .................................................. 7
      2.3.10 Change in OMXN40 .................................. 7
   2.4 Data Processing ............................................... 8

3 Results .............................................................. 9
   3.1 Residual Analysis of Full Model ............................ 10
      3.1.1 Normal QQ-plots ..................................... 11
      3.1.2 Linearity .............................................. 11
      3.1.3 Cook's Distance ...................................... 12
      3.1.4 Variance Inflation Factor (VIF) ....................... 13
      3.1.5 Correlation .......................................... 14
      3.1.6 Variable Selection: Forward Selection ............... 14
      3.1.7 Variable Selection: Backward Elimination .......... 16
      3.1.8 Cross Validation .................................... 17
   3.2 Reduced Model ................................................. 17

4 Discussion .......................................................... 19
   4.1 Analysis of the results .................................... 19
   4.2 Asymmetric Information ..................................... 20
   4.3 Unintentional Underpricing ................................ 21
   4.4 Media Coverage ............................................. 21
   4.5 Financial Performance Indicators ......................... 21
1 Introduction

1.1 Background

An initial public offering, or IPO, is the process of offering shares of a private company for the first time on a publicly traded stock market. When a company’s shares are made available for the public to trade, they are exposed to a bigger circle of potential investors. Undergoing an IPO is therefore usually done in order to raise more equity and thus improve their opportunities to expand. However, becoming a publicly traded company also entails higher requirements on documentation, the company’s disclosing of their finances and other business information compared to being private. Furthermore, the process of undergoing an IPO is both expensive and time consuming. The process consists of several phases including valuation of the company, due diligence, legal reviews, pre-marketing of the company etc. Consequently, undergoing an IPO involves many external actors. An extensive part of the process is referred to as the roadshow, in which the company is promoted to potential investors by the underwriters for the IPO and management from the company [1]. The roadshow requires a lot of time and money, however, as a consequence of the pandemic the process regarding the roadshow has been reworked in many countries. In contrast to travelling around the world meeting with various potential investors, most of the promoting part is now executed digitally [2].

Despite the economic strains in recent years, including a worldwide pandemic and a war in Europe, the market has experienced a high demand for IPO listings, and IPO underpricing has remained a prominent phenomenon. Based on an article published on Nasdaq by Phil Mackintosh, there has been a general trend of increasing average first day returns over decades. According to the article, the average increase of first day returns for IPOs from 1980 to 2020 was 18.4% [3]. This persistent trend has raised questions about the possibility of predicting first day returns and identifying the parameters that influence the change in share price during the initial trading day after an IPO.

1.2 Aim of the study

The study will be concluded through a multiple linear regression and aims to identify if there are any parameters, within a prechosen set of relevant parameters, that have a significant effect on the development of the share price during the first day after an IPO. The purpose of identifying a correlation between parameters and the share price is to generate a model that could explain and predict the first day return. Moreover, the study will also analyze whether the changes made in the process regarding an IPO, which arose as a consequence of the pandemic, has an impact on the first day return. Finding and creating such a model would be of interest for potential IPOs as well as private and institutional investors.
1.3 Research Question and Hypothesis

Taking into account the background and the aim of the study, following research questions have been formulated for the study:

- Does any of the preset variables have a significant relationship with the change in stock price during the first day after an IPO? Furthermore, is it possible to prove significant differences between the first day return for listings made previous to and post Covid?

- Given the predefined set of variables, are any of them more important for market participants to consider?

Considering the structural complexity of the financial market and the amount of unknown and immeasurable parameters influencing its movements, the hypothesis for the study is that the model will not be statistically significant. However, some variables are believed to explain some of the response variable’s variation.

1.4 Scope and Limitations

In order to enable an accurate and a precise regression model some limitations with the purpose to reduce the scope of the research were made.

Firstly, the study was narrowed down to only incorporate a set of Nordic market places:

- Nasdaq Stockholm
- First North
- NGM SME
- Spotlight

Because of differences in for instance regulations and structures between marketplaces globally, a geographical restraint was considered to be necessary. The chosen market places were assumed to be compatible for comparison with each other and together they include an adequate amount of data to be able to perform an accurate analysis.

Secondly, in order to obtain a reasonable and relevant set of data, the time frame of matter for the study was chosen to be listings made within the period of 2017-01-09 until 2022-12-09. During these years 418 IPOs were carried out on the market places specified above [4]. Furthermore, the chosen time frame also enables the project to analyze potential impact the pandemic might have had on the development of stock prices during the first day after an IPO.

Finally, it is widely acknowledged that identifying all the factors impacting
the first day return is nearly impossible [5]. Therefore, for the purpose of this study, it was essential to limit the number of explanatory variables included in the analysis. The variables included in the different models, namely the full- and reduced model, is specified in section 2.

1.5 Previous Research

As previously stated, IPOs have been a widely discussed and extensively researched topic over the years. Of particular interest is the observed pattern of initial positive first day returns, followed by long-term underperformance in IPOs, which has been thoroughly examined by researchers. M. Banu Durukan analyzes this aspect in her article The Relationship Between IPO Returns and Factors Influencing IPO Performance: Case of Istanbul Stock Exchange. The study found that the variables type of investors purchasing the shares and P/E ratio had an insignificant relationship with IPO returns, both in the short- and long-term. However the study found that the variables Debt-equity ratio and company age had positive betas, which indicated that a higher level of debt and an older company had a positive effect on the short-term returns. The negative betas of firm size, gross proceeds and method of IPO implies that shares of smaller firms, firms whose gross proceeds are low and IPOs that are sale of shareholder provide high short-term returns. Simultaneously the study also presented that bigger firms with a lower ownership retention generate lower long-term returns while shares of a privatized firms tend to perform better in the long run. In conclusion the findings in the study provide evidence that overevaluation of IPOs by investors as well as deliberate underpricing to reward the informed investors, are key factors in explaining the IPOs abnormal initial returns. Simultaneously the factors that showed a decrease in the uncertainty associated with the IPOs also showed lower returns. [6]

Another well researched aspect of IPOs is the phenomena of underpricing of IPOs. The article Initial Performance of IPOs in India: Evidence From 2010-2014, written by Sweety Nishant Shah and Disha Harshadbai Mehta, analyzes this aspect by examining the listings day performance while also analyzing parameters affecting the underpricing of IPOs on the Indian stock market. The study showed that neither issue price, issue size or market index return had a significant relation with the level of underpricing of the IPO. However, the variable overpricing appeared to have a significant linear relationship with the level of underpricing and hence also the first day returns for the IPO. A higher oversubscription denotes a higher demand of the issue which results in a higher first day return for the IPO. [7]
2 Method

2.1 Multiple Linear Regression
In this study a multiple linear regression model is utilized to elucidate the variation in the response variable by using a set of predictor variables, commonly referred to as the independent variables. The relationship between the response variable $y$ and the regressors $x_i$ is illustrated through below equation:

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_{18} x_{18} + \epsilon$$

The model relied on the following assumptions to be satisfied [8].

- The error has a mean of zero
- The error has a constant variance of $\sigma^2$
- The error terms are uncorrelated
- The error term is normally distributed with variance $\sigma^2$
- Linear relationship between the response variable and the regressors

A general model was generated using all the predefined regressor variables. To fulfill the assumption made about the error term being normally distributed a winsor transformation was made. The method limits extreme values in the data by setting them equal to the value of a specified percentile. This was done with the purpose of reducing their effect on the model. When applying winsorizing, the top and bottom 5 percent of the data was replaced. This model was then evaluated and analyzed through residual analysis, analysis of multicollinearity and cross validation. Additionally, when creating a regression model, data points will influence the model differently. Some data points were identified as outliers, having significantly large influence on the model, and they were investigated further. Based on the results from these analyses a new reduced model was generated and evaluated.

2.2 Response Variable
The aim of the model is to establish a relationship between the response variable and the predictors, allowing for an understanding of how changes in the independent variables impacts the dependent variable. [9].

2.2.1 Development in Stock Price During the First Day After an IPO
The variable the model aims to predict is the price development of the share price during the first day after an IPO. It can be calculated as the initial share price divided by the closing share price of the first day.
2.3 Regressor Variable

The regressor variables are the parameters included in the regression model with the aim to predict the response variable. In the study, 16 regressors were used, both dummy- and continuous variables.

2.3.1 Date of IPO

In order to capture potential differences in the price development during the first day after an IPO, previous and post the pandemic, a dummy variable presenting weather the listing was made during or after Covid was incorporated. According to Krisinformation.se the first Covid patient in Sweden was identified 24th of January 2020 [10]. Hence, listings made prior to that date were given a value of zero unlike the others which were given a value of one.

2.3.2 Marketplace

There are plenty of different marketplaces on which companies can be listed. Depending on for instance the size and structure and plan of the company, some marketplaces are better suited for the listing than others. The response variable was assumed to depend on the character and structure of a company. Taking this into account, the marketplace subject to an IPO is considered relevant to include in the model.

Nasdaq Stockholm is a regulated marketplace and it is governed by the Swedish Securities Market act [11]. It is the primary marketplace in Sweden and mainly has three types of listings:

- Nasdaq Small - Companies with a market capitalization under 150 million EURO
- Nasdaq Mid - Companies with a market capitalization between 150 million EURO and 1 billion EURO
- Nasdaq Large - Companies with a market capitalization over 1 billion EURO

First North is a Multilateral Trading Facility (MTF), i.e. it is not regulated, governed by Nasdaq Stockholm. The marketplace execute listing of smaller Nordic growth companies with the purpose of accessing growth capital. As of today First North has more than 540 listings [12].

Spotlight marketplace is in many aspects similar to First North. It is a MTF containing listings of smaller growth companies with the aim of simplifying the growth through accessing of capital. Spotlight has approximately 170 listings today [13].

NGM SME is an, by Finansinspektion, approved regulated Swedish marketplace. NGM SME offers listings and trades of various types of securities. As of
2023 there were 89 companies listed at the marketplace [14].

Nasdaq Small, Nasdaq Mid, Nasdaq Large, First North, Spotlight and NGM SME were all implemented as binary dummy regressors in the model.

2.3.3 Advisor
In the pursuance of an IPO, companies in general work with underwriters. They work closely with the company and their main job is to valuate the shares by buying them from the company and selling them to investors [15]. Since the advisors are key factors in the valuation of the company’s stock it is of interest to examine whether any advisors are more prone to either under- or overpricing a company’s stock.

2.3.4 Offering
Offering in an IPO is defined as the total value of the shares being offered to investors. The offering for each listing was measured in million SEK. A large IPO usually attracts attention in form of, for instance, a wide media coverage and a lot of mentions in social medias. The ”Offering” variable is assumed to correlate with the size of the IPO and hence explain some of the variation in first day returns.

2.3.5 Amount Newly Issued
When doing an IPO the company can both offer already existing shares and/or issue new ones for investors to buy. The issuing of new shares allows for the occurrence of adverse selection for the public investors[16]. Hence, some investors will perceive the issuing of new shares in an IPO as an implication of more risk. The variable is thus included with the purpose of capturing parts of a potential aversion towards investing in a specific IPO.

2.3.6 Initial Share Price
The initial share price refers to the introductory price at which the shares will be offered at to the public investors. This is predetermined by the underwriters and company management before the IPO and is based on the valuation of the company combined with the amount of shares being offered [17]. A low initial share price allows a greater proportion of potential investors to buy the share. Thus, the variable is assumed to capture the availability of buying the shares and hence also describe some of the variation in the response variable. The initial share price was measured in amount of million SEK.

2.3.7 Market Capitalization
Market capitalization refers to the total value of all stocks, calculated by multiplying the number of stocks with its price. In our model, the market cap-
italization value at the time of the IPO was used. Similar to the "Offering" variable, this regressor is presumed to capture the level of hype or excitement surrounding the IPO.

2.3.8 Amount of Shares Pre Subscribed

The process of an IPO usually includes a roadshow which allows some investors to buy shares before the actual listing. This variable refers to the percentage of shares bought before the shares were available for the public to purchase. If the amount of shares being presubscribed is large, it is assumed to increase the demand of the shares and thus clarify some of the variance in the response variable.

2.3.9 SPAC

A Special Purpose Acquisition Company (SPAC) is a non operating company which is founded with the ambition of acquiring a non listed company with the money gathered from the selling of shares \[18\]. When the acquisition is made the purchased company also becomes listed. Thus, SPACs enables listings without the bought company having to go through a traditional IPO process.

Since a SPAC in the beginning does not have any actual value, the initial share price for a SPAC being listed is usually fixed. At Swedish market places this price is normally 100 SEK \[19\]. Given the analyzed marketplaces and, the model had to consider seven listed SPACs. The variable is implemented as a binary dummy variable, taking the value one if the listing was a SPAC and zero otherwise. Since SPACs differs in fundamental ways to a traditional company, it was assumed have an impact on the response variable.

2.3.10 Change in OMXN40

OMXN40, or OMX Nordic 40, is a stock market index consisting of the 40 largest and most traded companies on the Swedish, Finish and Danish stock markets. The purpose of using this stock market index in the model was to measure stock markets performance at the time of the IPO and hence capture some of the general investor risk assessment in Europe.
2.4 Data Processing

The main data set used for this study was obtained from Affärsvärlden. Data about the development in OMXN40 was obtained from Nasdaq’s website.

The first step in processing the data was removing all IPOs that had no data available for share price development during the first day. Most of these data points were IPOs that weren’t carried through. These were marked as cancelled in the original data set, or IPOs that had not yet been offered on the market.

Secondly all the IPOs listed on Nasdaq Stockholm in the data set were split into Nasdaq Small, Mid or Large depending on the requirements for the company’s market capitalization mentioned earlier. The only company in the data set listed on Nasdaq Helsingfors was also assigned to either Nasdaq Small, Mid or large depending on its market capitalization. It was assumed that Nasdaq Stockholm and Helsingfors were very similar marketplaces since both are regulated markets governed by Nasdaq. The same assumptions about similar marketplaces and regulations were made about First North and First North Denmark and therefore all IPOs listed on First North Denmark were assigned to First North. The remaining marketplaces, Oslo Børs and Merkur Market, both contained only a couple of IPOs and were therefore removed from the data set used to generate the model.

Thirdly the advisors were categorized into four different groups. Carnegie, Sedermera, Eminova and OtherAdvisor since these three investment banks advised the most IPOs. Any IPO that had several different advisors where Carnegie was one of them were assigned to Carnegie. In cases where non of the above mentioned advisors were involved they were assigned to the variable OtherAdvisor.

Finally one IPO was removed from the data set due to no data of the OMXN40 development for that particular date.
3 Results

The following values for each variable were retrieved from the full regression model:

| Coefficients: (2 not defined because of singularities) | Estimate | Std. Error | t value | Pr(>|t|) |
|------------------------------------------------------|----------|------------|---------|---------|
| (Intercept)                                           | 9.299e-02| 1.632e-01  | 0.570   | 0.5691  |
| Date                                                 | 3.292e-02| 2.925e-02  | 1.126   | 0.2610  |
| Offering                                              | 1.318e-05| 2.723e-05  | 0.484   | 0.6287  |
| AmountNewlyIssued                                     | -7.827e-02| 5.939e-02 | -1.318  | 0.1883  |
| InitialSharePrice                                     | 4.103e-04| 6.442e-04  | 0.637   | 0.5246  |
| MarketCap                                             | -2.235e-06| 3.344e-06 | -0.668  | 0.5044  |
| PreSubscribed                                         | 1.456e-01| 5.948e-02  | 2.448   | 0.0148 *|
| SPAC                                                  | -1.199e-01| 1.158e-01 | -1.035  | 0.3011  |
| OMX40                                                 | -1.218e+00| 1.673e+00 | -0.728  | 0.4670  |
| Spotlight                                              | -2.411e-02| 1.581e-01 | -0.152  | 0.8789  |
| FirstNorth                                            | -8.631e-02| 1.524e-01 | -0.566  | 0.5716  |
| NGMSME                                                | -2.410e-01| 1.614e-01 | -1.492  | 0.1364  |
| NASDAQSmall                                           | -6.278e-02| 1.615e-01 | -0.389  | 0.6977  |
| NASDAQMid                                             | -7.121e-03| 1.416e-01 | -0.500  | 0.6099  |
| NASDAGLarge                                           | NA       | NA         | NA      | NA      |
| Carnegie                                              | 1.690e-02| 5.966e-02  | 0.283   | 0.7771  |
| Sedermera                                             | 4.608e-03| 5.014e-02  | 0.902   | 0.3628  |
| Eminova                                               | -5.441e-02| 5.135e-02 | -1.080  | 0.2899  |
| OtherAdvisor                                           | NA       | NA         | NA      | NA      |

---

Signif. codes:  
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2793 on 394 degrees of freedom
Multiple R-squared: 0.09172,  Adjusted R-squared: 0.05483
F-statistic: 2.487 on 16 and 394 DF,  p-value: 0.001255

Figure 1: Table of the coefficients

As illustrated in Figure 1 the variables NASDAQLarge and OtherAdvisor were not given any values, instead it said NA. This occurs because of singularity, i.e. the variables of question had a perfect linear relationship with other variables. These variables were assumed to not contribute to the model and were therefore disregarded. Furthermore, the model generated a low adjusted $R^2$ of 0.05483. This suggests that the model only was able to explain approximately 5.5% of the variation in the response variable. Additionally, PreSubscribed was the only variable which had a significant p-value.
3.1 Residual Analysis of Full Model

In order to proceed an analysis of the normal- and standardized residuals, the residuals were plotted for each data point.

![Graphs illustrating the Normal and Standardized residuals](image)

From the left graph in Figure 2 it can be identified that the residuals are centered around a value of zero, with the exception of a few points. Additionally, the sum of the residuals was calculated to $1.639314 \times 10^{-16}$. Hence, the assumption about the error term having a mean of zero is considered to be sufficiently accurate.

In the graph to the right in Figure 2, illustrating the standardized residuals in terms of standard deviations, it can be seen that none of the points breaches the often used cutoff of 3 standard deviations [8]. However, several residuals are close to the cutoff, which indicates that there might exist some outliers. Considering that the windsorizing replaced the extreme values of the predicted variable, the potential outliers probably differ from other points in regards to some of the predicting variables.

![Graph of the residuals in relation to the fitted values](image)

From the graph in figure 3 it can be identified that the residuals seem to have a faintly double bow pattern, however, in general the pattern seems to be rel-
ative random. Hence, the assumptions that the error terms are uncorrelated and have constant variance can be assumed to be satisfied.

### 3.1.1 Normal QQ-plots

![Normal QQ-plot of the residuals before and after Winsorizing](image)

Figure 4: Normal Q-Q plot of the residuals before and after Winsorizing

Looking at the left graph in Figure 4 it can be seen that the majority of the residuals previous to the transformation follow the pattern of a normal distribution. However, as seen in the graph beginning from approximately quantile 2, the data starts to deviate from the normal distribution. Specifically we observe that the points 136, 232 and 271 deviate significantly. Through further analysis of the data it was identified that these points had remarkably large development in share price during the first day after the IPO (see Safello, QuickBit and Stenocare in excel). Instead looking at the right graph in Figure 4, the assumption about the error term being normally distributed after winsorizing is considered to be valid.

### 3.1.2 Linearity

The relationship between the response variable and a majority of the regressors can not be analyzed since most of the regressors are binary dummy variables. The remaining regressors were plotted against the response variable to examine for linear relationship. The x-axis was scaled down for the regressors Offering, InitialSharePrice and MarketCap due to a wide spread between the values of a majority of data points and a few very large values.
As seen in the graphs above there are no clear signs for any of the variables having an obvious linear relationship with the response variable. However, some linear patterns can be identified for AmounNewlyIssued and PreSubscribed. Hence, the assumption about the response variable having a linear relationship with the regressors is considered to only be satisfied for those two variables.

### 3.1.3 Cook's Distance

In order to further analyze potential outliers, the Cook’s distance for each point was plotted.

The general cutoff in Cook’s distance for a point to be considered as an outlier is 1 [8]. Looking at the graphs in Figure 6 it was concluded that none of the data points, according to the theory behind Cook’s distance, should be treated as a
high influential outlier. However, it is possible to identify point 232 (QuickBit) as a point having large influence on the model, in comparison to others.

3.1.4 Variance Inflation Factor (VIF)

The Variance Inflation Factor, or VIF, is a measurement indicating the combined effect of the dependencies among the regressors on the variance of that term. Hence it can be used to determine Multicollinearity in the model. According to Montgomery et al. (2012) a VIF value greater than 5 or 10 indicates a poorly estimated Beta and should therefore be considered to perhaps be removed from the model.

The VIF values for each regressor in the model is presented in the table below.

<table>
<thead>
<tr>
<th>Regressor</th>
<th>VIF-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date</td>
<td>1.1253</td>
</tr>
<tr>
<td>Offering</td>
<td>11.6290</td>
</tr>
<tr>
<td>Amount Newly Issued</td>
<td>1.6735</td>
</tr>
<tr>
<td>Initial Share Price</td>
<td>2.0529</td>
</tr>
<tr>
<td>Market Cap</td>
<td>8.4995</td>
</tr>
<tr>
<td>Pre Subscribed</td>
<td>1.1722</td>
</tr>
<tr>
<td>SPAC</td>
<td>1.1828</td>
</tr>
<tr>
<td>OMXN40</td>
<td>1.0342</td>
</tr>
<tr>
<td>Spotlight</td>
<td>22.3440</td>
</tr>
<tr>
<td>First North</td>
<td>30.4550</td>
</tr>
<tr>
<td>NGMSME</td>
<td>11.5210</td>
</tr>
<tr>
<td>NASDAQ Small</td>
<td>5.7558</td>
</tr>
<tr>
<td>NASDAQ Mid</td>
<td>8.4437</td>
</tr>
<tr>
<td>Carnegie</td>
<td>2.0720</td>
</tr>
<tr>
<td>Sedermera</td>
<td>1.3909</td>
</tr>
<tr>
<td>Eminova</td>
<td>1.1100</td>
</tr>
</tbody>
</table>

Table 1: Table containing the VIF values for all the regressors in the full model

Based on the table above it is clear that several of the variables have a VIF value exceeding the value 10 and was therefore further investigated for the design of the reduced model.
3.1.5 Correlation

To further analyze the Multicollinearity in the model and the high VIF values a correlation plot was plotted.

![Correlation Plot](image)

Figure 7: Graph illustrating the correlations between each variable in the data set

From the plot above it can clearly be seen that MarketCap and Offering are very heavily correlated. It can also be identified that the market places First North and Spotlight are quite correlated. Other variables that show a higher correlation with several other variables are InitialSharePrice, Carnegie and Amount-NewlyIssued.

3.1.6 Variable Selection: Forward Selection

In order to find a optimal subsets of regressors for the model, forward selection was made.
The Adjusted $R^2$ is a measurement indicating how well the data fits the regression line while also adjusting for the number of variables in the model. According to Montgomery et al. the Adjusted $R^2$ decreases when variables not contributing too capturing variance in the response variable are included, and increases when relevant variables for the model are added [20]. Therefore the criterion for selecting the optimal subset of regressors considering the adjusted $R^2$ model is to find the maximum of the adjusted $R^2$.

Based on the graph on the top left of Figure 8 the highest adjusted $R^2$ value is obtained with six regressor variables (marked with red in figure 9). However it can also be noted that a model containing four, five, seven or eight regressors generate similar adjusted $R^2$ values. The top right graph in Figure 8 illustrates these optimal sets of variables. For six variables in the model the optimal subset of regressor variables are: Date, AmountNewlyIssued, PreSubscribed, First North, NGM SME and Eminova. For five variables the optimal subset contains the same variables as for the model with six regressors, except Date. And the optimal subset for four variables in the model are the same variables as for the model with six regressors minus the variables Date and Eminova.

The Mallows $C_p$ Statistic metric was used to determine the bias of introducing different regressors into the model. Mallows $C_p$ Statistic was calculated by dividing the residual sum of squares $SS_{RES}$ by the mean squared error (MSE), which both preferably should be minimized to obtain a good model [20]. Hence we also want to minimize the Mallows $C_p$ Statistic.
From the bottom left graph in figure 8 we can conclude that the amount of regressor variables in the model which minimizes Mallows $C_p$ is four. Again these four variables can be identified from the bottom right graph in figure 8. Hence, the optimal subset according to Mallows $C_p$, using forward selection, for the model is AmountNewlyIssued, PreSubscribed, First North och NGM SME.

### 3.1.7 Variable Selection: Backward Elimination

Similarly to forward selection, backward selection was made with the purpose to find an optimal variable set for the model.

![Adjusted $R^2$ and Mallows $C_p$ Statistic generated by Backward Selection](image)

Based on the top left and right graphs in Figure 9 we can conclude that the maximum Adjusted $R^2$ value, using backward elimination, is just like in the forward selection obtained with a model containing six regressors. It can also be identified that the optimal subset for this number of variables are the same as for the forward selection.

Just like the Adjusted $R^2$ value the Mallows $C_p$ Statistic is minimized by the same number of the regressors as in the forward selection. It can also again be concluded that the optimal regressors for this subset of four variables are the same as in the forward selection.
3.1.8 Cross Validation

Cross Validation was used to estimate the accuracy of the predictive performance of the model, where a lower value indicates that the prediction is good. The graph below illustrates the Cross Validation values for models containing different number of regressors.

![Cross Validation Graph](image)

Figure 10: Graph illustrating the Cross Validation value for each number of variables in the model

Based on the graph above it can be seen that the amount of regressor variables in the model that generate the lowest cross validation values are four or two regressors. Four regressors give a slightly lower value than two. These four variables and their cross validation value is presented in the table below.

<table>
<thead>
<tr>
<th>Regressor</th>
<th>AmountNewlyIssued</th>
<th>PreSubscribed</th>
<th>FirstNorth</th>
<th>NGMSME</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>-0.12676071</td>
<td>0.14259277</td>
<td>-0.06023635</td>
<td>-0.23129033</td>
</tr>
</tbody>
</table>

Table 2. Table containing the optimal set of variables and their coefficient values obtained from cross validation

From the information in the table above it can be noted that the regressor variables that generate the lowest Cross Validation value are the same variables as the ones that were obtained using variable selection.

3.2 Reduced Model

Based on the results and analysis of the full model the regressors chosen for the reduced model were AmountNewlyIssued, PreSubscribed, FirstNorth and NGM SME. The following values for each variable were retrieved from the reduced regression model.
As can be seen in the table above a higher adjusted $R^2$ value and a lower p-value was obtained for the reduced model. Unlike the full model, all regressors in the reduced model seemingly had a statistical significant relationship with the response variable. This is illustrated through p values within a 95% confidence interval, i.e. p values lower than 0.05, for all included regressors.

The assumptions made about the error term and relationship between the response variable and the regressors need to be checked for the reduced model.

Figure 11: Table of the coefficients for the reduced model

| Coefficients   | Estimate | Std. Error | t value | Pr(>|t|) |
|----------------|----------|------------|---------|----------|
| (Intercept)    | 0.13878  | 0.04870    | 2.850   | 0.00460  ** |
| AmountNewlyIssued | -0.12676 | 0.04739    | -2.675  | 0.00778  ** |
| PreSubscribed  | 0.14259  | 0.05702    | 2.501   | 0.01279  * |
| FirstNorth     | -0.06024 | 0.03035    | -1.985  | 0.04784  * |
| NGMSME         | -0.23129 | 0.05156    | -4.486  | 9.47e-06  *** |

---

Signif. codes:  * '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2771 on 406 degrees of freedom
Multiple R-squared:  0.0786,  Adjusted R-squared:  0.06952
F-statistic: 8.659 on 4 and 406 DF,  p-value: 1.028e-06

Figure 12: Graphs illustrating the residuals, residuals in relation to the fitted values and a Normal Q-Q plot of the residuals
Firstly it can be assumed from the graph in the upper left corner of figure 12, and the calculated sum of residuals being $7.866971e-16$, that the assumption about the mean of the error term being 0 is satisfied. Secondly since the graph in the upper right corner show no particular pattern and the residuals seem relatively randomly distributed it can be concluded that the assumptions about uncorrelated error terms and constant variance is adequately accurate. Thirdly the normal Q-Q plot in the bottom left corner indicate that the normal distribution assumption is fulfilled. Finally, both the variables Amount Newly Issued and Pre Subscribed were the variables that showed the most tendencies of a linear relationship with the response variable in the full model. First North and NGM SME are both binary dummy variables, hence this assumptions can not be validated for these variables.

Both First North and NGM SME showed signs of multicollinearity in the full model due to high VIF values. The reduced model generated the following VIF values.

<table>
<thead>
<tr>
<th>Regressor</th>
<th>VIF-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amount Newly Issued</td>
<td>1.0824</td>
</tr>
<tr>
<td>Pre Subscribed</td>
<td>1.0946</td>
</tr>
<tr>
<td>First North</td>
<td>1.2260</td>
</tr>
<tr>
<td>NGM SME</td>
<td>1.1937</td>
</tr>
</tbody>
</table>

Table 3. Table containing the VIF values for all regressors in the reduced model

The values in the table above show low VIF values for all regressors in the reduced model, hence all regressors are assumed to be sufficiently independent and not explained by other variables in the model.

4 Discussion

4.1 Analysis of the results

The presented result shows that neither of the models are good at predicting the outcome of first day return. The full model only explains about 5.5% of the variation in the response variable, according to the $R^2$ value. The poor performance is based on several reasons. Firstly, it can be seen that none of the chosen regressors have a clear linear relationship with the response variable. Since the method used for the creation of the model was multiple linear regression, the lack of linearity obviously impacted the models accuracy negatively. Secondly, when analyzing the model for multicollinearity, several of the variables were showed to correlate with each other. This implies that the prechosen set of explanatory variables was not optimal in regards to explaining the variance of the response variable. Furthermore, the existence of multicollinearity can heavily impact the precision of the regression coefficients and thus the reliability of the model [8]. Thirdly, the chosen time frame of the study limits the amount of IPOs for the model to train on. Few observed data points increases the risk of generating an overfitted model. That is, the model would predict future first day returns of IPO inaccurately. Lastly, it was known from before that the structural complexity trailing the fluctuations on the stock market would make it difficult to find and measure all the relevant factors influencing the response variable. The bad performance of the model indicates the lack of relevant explanatory variables.
The reduced model was created with the purpose of optimizing our model. From
the result it can be identified that it is better at explaining the variance in the re-
response variable, having an adjusted $R^2$ corresponding to approximately 7% of the
variance. Additionally, the VIF values of the reduced model indicates that the high
correlations that were identified in the full model have been adjusted for by removing
other correlating variables. The variables included in the reduced model were cho-
sen from the variable selection made of the full model. The result illustrated that
NGMSME explains most of the variance in the response variable out of the predefined
set of variables, and its coefficient indicates NGMSME having a negative impact on
the first day return. This result probably originates from the fact that the analyzed
data only included 38 IPOs made on the specific marketplace, and a large majority
of them experienced a negative share development during the first day as a publicly
traded company. PreSubscribed showed a positive relation with the response vari-
able, suggesting that the greater the percentage of the offer that is pre subscribed,
the greater the first day return. This result is in line with the hypothesis stating that
the amount of pre subscribed shares increases the demand of the shares. Moreover,
AmountNewlyIssued had a negative impact on the response variable, stipulating that
a higher percentage of the offered share being newly issued corresponds to a lower
demand.

To further evaluate the reduced model the other optimal variables suggested based
on the adjusted $R^2$, Eminova and Date, were added. Both a model containing five
regressors and six regressors, gave almost exactly the same adjusted $R^2$ value as the
model with four regressors. However, neither Eminova or Date showed a statistical
significant relationship with the response variable when added to the model. Hence,
it was decided to keep the reduced model consisting of only the four variables listed
earlier.

For an accurate and precise modelling, it is essential to include all parameters that ex-
plain the variance in the response variable. Looking at the performance of the created
models, it can be concluded that the chosen variables fail to capture enough variance
in order for them to be considered precise models.

4.2 Asymmetric Information

In line with the Winner’s theory, created by Rock, the demand for a share, and thus the
price of it, depends on the information the investors possesses. Investors are usually
divided into two groups of which the first one contains informed investors and the
second of uninformed investors. On one hand, since the first group possesses superior
information they tend to only invest in financially lucrative IPOs, often underpriced
listings. On the other hand, the second group will in general invest in both the
under- and overpriced IPOs, due to lack of information [21]. Hence, the demand
will differ depending on the size of respective group which onwards will impact the
price development of the share. Capturing asymmetric information as a variable was
assumed to be almost impossible, and thus the models were not able to include this.
4.3 Unintentional Underpricing

In some cases there occur different perceptions of the valuation of the company which arises from uncertainties regarding the intrinsic value, even in situations where all inventors are exposed to the same information [22]. This might lead to a certain degree of unintentional underpricing from the underwriters which also tends to increase in magnitude the more uncertainty there is. Consequently, the share price after the IPO will be affected by the different perceptions of the value of the company. However, including unintentional underpricing as a factor in the model was assumed to be too complex.

4.4 Media Coverage

The media plays a significant role in reducing the information asymmetry, often referred to as signaling theory, by providing information to investors and increasing the transparency of the company. Transparency can both enhance the company’s status and help the company gain legitimacy with the investors [23]. Normally, mentions in various medias imply a better general knowledge about the IPO, hence decreasing the risk for asymmetric information. However, measuring media coverage is not as simple as counting the number of mentions in various forums. Not all information published on different media forums are equally valuable. Coverage in a credible financial source has far more influence on the investor’s investment decision than for example a twitter post. Moreover, to capture the influence of media mentions one has to account for whether the post portrays the IPO in a positive or negative manner. That is, will the publication lead to more or fewer investments. Since investments in IPOs often are considered as more risky and uncertain than already established and listed companies, the tone in media has a significant impact on the investor’s perception of the company. Against the background of this, the models were not able to include media coverage as a variable.

4.5 Financial Performance Indicators

Financial performance indicators, such as Price-to-Earnings (P/E) ratio and Debt-Equity ratio, are additional factors that potentially would have had improved the performance of the model. The P/E ratio is a measurement of the current share price in relation to the earnings per share and it is often used to compare a company’s relative value against similar companies within the same industry. Besides being a relevant measure to compare different companies stock values it also gives an indication whether the stock is overvalued or undervalued. For instance, a higher P/E value could potentially indicate that the expectations on the stock performance from investors are high, or that the stock is overvalued.[24] Hence, it could have been a relevant variable to include in the model, potentially explaining some of the variance caused by expectations on the IPO.

The Debt-Equity ratio is a measurement of a company’s financial leverage. It shows how much of the operation which is financed through debt rather than its own equity. Debt financing implies greater distress costs and larger vulnerability towards increased rates. Hence, Debt-Equity ratio gives an indication on the risk of investing in the company.[25]
To measure these factors one would have to manually search through each of the over 400 companies’ annual reports to gather the relevant financial data. The time this would have taken became a limiting factor for this study, hence it was not included in the model. However, including financial performance indicators, such as P/E and Debt-Equity ratio, would have added some aspects to the model that were not captured in the study. Other variables that could possibly also be included in any further studies to capture the risk and expectations are for instance the age and size of the company issuing the IPO.

5 Conclusion

In conclusion, the models created in the study were not able to predict the first day return of an IPO accurately. By looking at the adjusted $R^2$ value for the two models it can be identified that their ability to explain the variance in the response variable is too poor for them to be considered reliable models. However, given the predefined set of variables used, some of the variables were showed to explain the change in first day return better than others. NGMSME, First North, PreSubscribed and AmountNewlyIssued all showed statistical significance with the response variable in the reduced model.

Additionally, no signs of the pandemic having an impact on the first day return were found. Hence, the hypothesis about that the changes in the process regarding an IPO, which arose from the suites of the pandemic, would have an impact on the first day return is according to this study false.

6 Further Studies

To generate a better performing and more accurate model, further studies on the subject should consider including variables capturing for instance financial performance indicators, investor’s expectations and media coverage. As discussed in section 4, such variables have been shown by several studies to have an impact on the response variable. By failing to adjust for all relevant explanatory variables, the model runs the risk of, among other things, omitted variable bias.

Moreover, the analyzed marketplaces and chosen time frame in the study limited the amount of available data. Expanding the scope of the study by using a bigger data set would provide a more reliable estimate of the underlying statistical significance between the included variables. Thus, more data implies a larger likelihood of the model being able to capture the true underlying patterns and relationship in the data. Additionally, by accounting for more marketplaces and/or analyze IPOs over a longer time period would increase the robustness of the model and decrease the risk for overfitting.

In order to further develop on the subject it could be of interest to also investigate the post-IPO performance of companies. This could provide valuable insights into the long-term sustainability and profitability of IPOs. Examining variables such as financial performance and market valuation over a longer period of time could shed light on whether positive or negative first day returns persists in the long run. Such research has the potential of providing investors, policymakers and other market par-
ticipants with better knowledge about the risks and rewards associated with initial IPO investments and contribute with relevant implications for future IPO strategies and decision making.
References


