The Impact of Hype on IPO First-Day Returns

A study of IPOs on the exchanges operated by Nasdaq Stockholm

EMIL ISAKSSON

MIKAEL KARPE
The Impact of Hype on IPO First-Day Returns

A study of IPOs on the exchanges operated by Nasdaq Stockholm

EMIL ISAKSSON
MIKAEL KARPE

Degree Project in Applied Mathematics and Industrial Economics (15 credits)
Degree Progr. in Industrial Engineering and Management (300 credits)
Royal Institute of Technology year 2016
Supervisors at KTH: Henrik Hult, Jonatan Freilich
Examiner: Henrik Hult

TRITA-MAT-K 2016:17
ISRN-KTH/MAT/K--16/17--SE

Royal Institute of Technology
SCI School of Engineering Sciences

KTH SCI
SE-100 44 Stockholm, Sweden

URL: www.kth.se/sci
Abstract

In the last two years, the number of Initial Public Offerings (IPOs) on the exchanges operated by Nasdaq Stockholm have been many. In parallel to this, the newly listed companies’ first-day returns have managed to outperform the market. In the meantime, the IPOs attain much attention that partly consists of general hype and rumors. As of this, it is of interest to investigate the correlation of hype and the first-day returns.

This study intends to identify the impact of hype statistically and analyze how the hype can be derived from media’s portraying. The analysis was performed with the help of regression analysis combined with a qualitative case study of media attention with the foundation from behavioral finance.

In short, the results of the analysis were statistically insignificant. Thus, it cannot verify that the hype has any impact on the first-day returns. In the meantime, the qualitative case study supports the hypothesis that hype would influence. As a result of this, further studies are necessary to draw clear conclusions concerning the impact of hype.
Sammanfattning


Den här studien hade som mål att statistiskt identifiera hur mycket hype påverkar förstadagsavkastningen och samtidigt visa hur medias framställning av börsnoteringen förklarar denna hype. Detta har utförts med statistiska medel som regressionsanalys samt med hjälp av en kvalitativ fallstudie av medias bevakning. I den kvalitativa studien har beteendeekonomi varit utgångspunkten.

Sammanfattningssvis lyckades inte studien statistiskt säkerställa att hype påverkar förstadagsavkastningen. Däremot stödjer den kvalitativa fallstudien hypotesen om att hype skulle ha inverkan. Slutsatser av resultaten blir därför svåra att dra och vidare studier krävs för att kunna konfirmera hypotesen.
4.2.1 Removal of Financial Data .................................. 19
4.3 Regression Model .............................................. 19
  4.3.1 Response Variable ........................................ 19
  4.3.2 Covariates ................................................. 20
  4.3.3 Base Model ................................................. 22

5 Results ............................................................ 23
  5.1 Linear Model .................................................. 23
  5.2 Log-transformation of Hype .................................. 25
  5.3 Log-transformation of First-Day Returns ................... 27

6 Discussion .......................................................... 30
  6.1 Discussion of Regression Results ............................ 30
    6.1.1 Implications From Residual and Q-Q Plot ............. 30
    6.1.2 Implications From Regressions ......................... 30
    6.1.3 The Linear Model ....................................... 30
    6.1.4 Log-transformation of Hype ............................ 31
    6.1.5 Log-transformation of First-Day Return ............... 31
  6.2 Sources of Errors ............................................ 32
    6.2.1 Mathematical Sources of Errors ....................... 32
    6.2.2 Complexity of Quantifying Hype ....................... 32
    6.2.3 Discounts in IPOs .................................. 33
    6.2.4 Missing Covariates .................................. 33
  6.3 Summary ...................................................... 33

7 A Case Study of the Media’s Portraying ....................... 35
  7.1 The Tobii Case ............................................. 35
  7.2 The Inwido Case ............................................ 36
  7.3 Summary ..................................................... 37
  7.4 Differences in the Two Studies ............................ 37

8 Conclusions ...................................................... 39

9 References ....................................................... 40

10 Appendix ........................................................ 45
List of Figures

1  A hypothetical value function ............................... 5
2  Heteroskedastic and homoskedastic residuals ............... 9
3  Plots for linear regression generated in R .................. 24
4  Plots for log-transformation of hype generated in R ........ 26
5  Plots for log-transformation of first-day return generated in R 29

List of Tables

1  Linear regression ........................................... 23
2  Confidence intervals ....................................... 23
3  VIF test .................................................. 23
4  Log-transformation of hype ................................ 25
5  Confidence intervals: Log-hype ............................. 25
6  VIF test: Log-hype ........................................ 25
7  Log-transformation of First-Day Returns ...................... 27
8  Confidence intervals: Log-return ............................ 28
9  VIF test: Log-return ..................................... 28
1 Introduction

1.1 Background

On the stock exchanges operated by Nasdaq Stockholm (Nasdaq OMX Nordic Stockholm and Nasdaq First North), there was a significant increase of IPOs during the years of 2014 and 2015. Also, the average IPO’s first-day return outperformed the markets. (Nilsson & Sandberg, 2016) The outperformance raises the question of what causes these first-day movements, and it is therefore of interest to understand what factors might be driving it.

An IPO is the first time a company is offered for trading on an authorized stock exchange and by this, any investor can access and own the stock. The reasons for a company to go public are several. However, the main reason is usually to attain capital and liquidity to reinvest in the company. Much more could be written on the reasons to perform an IPO. However, that is not part of the scope of this study.

During the IPO process, the investment bank value the company and later sells it, where the sale step is where the first-day returns will appear. This first-day return is commonly said to be driven by something called underpricing and the overall attention of the market towards the company.

Underpricing conducted by the underwriters is as mentioned one source of the on average high first-day returns of the IPOs and is explained in more detail in the next section. However, more interestingly, IPOs tend to draw much attention to them. The attention triggers irrational investing behaviors (Wermers 1999, 581–622) that might have an impact on the first-day return. The attention that a company attracts before their IPO will in this report be called hype.

In general, hype can in this context be seen as the total buzz around a certain company or some product (or similar) connected to the company. Hype is by this reason connected to rumors and speculations related to a company. Moreover, it is commonly said that hype affects the financial markets, however to what extent is unclear. As of this, the study is of interest as it will investigate how the hype impacts the market at the IPO stage.

In order to test the true impact of hype mathematically, an attempt to quantify hype is necessary. Google Keyword Tool measures the number of queries for a specific word during a period and is, therefore, a means to
attempt to quantify hype. Through regression analysis, the quantified hype can bring more precise knowledge of how the hype affects the first-day returns of IPOs in reality. To complement the statistical investigation, the area of behavioral finance will be used to explain further how the hype is generated and affects the first-day return.

1.2 The IPO Process

As previously mentioned, a company undergoes an IPO with several purposes. Normally, it is to attain money and capital for investments such as R&D, expansion or acquisitions. Other reasons might be that a company improves its international reputation or simply that the owners want to sell the company and that the IPO market is attractive for the moment. When a company decides to go public, a long process starts, where both the company have to meet certain standards in their business and operations. After the company has become mature for the IPO, the IPO process begins. (Bala Subramaniam, 2016)

The IPO process can be generalized with a few common steps. Normally, the company that is going public hires an investment bank or a financial institution (also referred to as an underwriter). The underwriter will be responsible for the whole process, and more importantly determining the correct price of the underlying company. The price is one of the most important parts to determine correct, as the bank needs to raise the sufficient amount of money to their client while making sure that the bank gets any profits.

In practice, this is done by letting the underwriter buy the stocks of the underlying company. As of this, the underwriter manages to raise sufficient money for the underlying company. Later on, the underwriter will sell their stocks to the originally intended investors for a higher price and hopefully make profits. However, as the underwriter does not intend to own the underlying company, it is important that the bank manages to sell as many stocks for an as high price as possible. When selling the stocks the underwriter tends to give a discount on their valuation and try to hype the stock. (Mark Koba, 2013)
1.3 Project Description

The project should find how hype affects the first-day return for IPOs on the exchanges operated by Nasdaq Stockholm and also investigate how media’s portraying of the new stock explains the hype. Therefore, the following two research questions have been developed:

1) What effect does hype have on IPO first-day returns from a statistical perspective?

2) How does the media’s portraying of an IPO explain the hype and the first-day returns from a behavioral finance perspective?

1.4 Aim of Study

The aim of the study is to try to generalize and quantify how hype affects the first-day return on a stock meanwhile qualitatively bringing an understanding of how the hype is derived from the media. As of this, the aim is to bring a clearer understanding for how the elements of behavioral finance and the actual financial outcome synchronizes with each other.

Furthermore, similar studies of explaining the effects of hype through statistics have been performed before. This research paper will hopefully help other to find inspiration and ideas to extend and advance the research in the field.
2 Theoretical Frameworks

2.1 Behavioral Finance

Behavioral finance, also referred to as behavioral economics, is the study concerning decision making in finance and economics, mainly based on theories from psychology. This field questions an essential assumption in economics and finance: the assumption of rationality and the assumption that decisions, in general, are taken to maximize the individuals' value. (Blume & Easley 2007, 1-15)

2.1.1 Prospect Theory

The prospect theory (also referred to as loss aversion) states that people normally tend to value losses and profits differently, where a loss (compared to an equally large profit) would be more negatively valued than the profit would have been positively valued. In practice, this would mean that a loss of $10 is valued more negatively than a profit of $10 is valued positively. In the context of finance, this means that investors should be more averse to a risk of losing than attracted to an equally large profit. Thus, this theory can help describing and explaining why investors act in certain ways in an IPO. (Tversky & Kahneman. 1979, 18-19)

The picture below represents how the prospect theory looks graphically. Notice that the slope is steeper on the right-hand side, and there is a diminishing effect on the loss aversion as well as the attraction for a profit. This means that a profit or loss adds less positive or negative value the larger the profit respectively loss becomes.

2.1.2 Decision Heuristics

A heuristic is any approach or method for solving a problem in order to make a decision. It is typical that the heuristic method serves to find a "good enough" or "fast enough" answer as the problem might be analytically complicated to solve. The heuristics normally works well for simpler purposes. However, it might sometimes make a person make misleading judgments and thus choose a non-optimal decision. (McNeill et al. 2012, 26)
Availability Heuristic
This heuristic is when the probability of an event or topic is overestimated because of a certain memory of a similar event or topic is readily available in the mind (Haiybor et al. 2008, 154). In the case of IPOs, if there have been many recent IPOs where the stock skyrocketed during the first trading day just before the upcoming IPO, an investor would be likely to overestimate the probability that the stock of the upcoming IPO also will skyrocket during the first trading day.

Anchoring
Anchoring is when an individual relies too much on the first benchmark they are exposed for, and thus their judgment of something around will be centered too heavily around that benchmark. Naturally, if a person has been exposed to this reference point, their judgment will likely be nearer the benchmark compared to if the individual would not have been exposed to it. It is commonly used in negotiations, and it is probable that this will have an effect on the stock price movement. (Tversky & Kahneman 1974, 1128-1129)

Affect Heuristic
The word 'Affect' is in this context a psychological term which describes how feelings or affections for something appear quickly and automatically in the mind. The heuristic is better explained by an example, where the words treasure and lung cancer generates two completely different feelings around them. When people are victims of this while assessing risks and benefits, this is called the affect heuristic. This heuristic could explain what makes people take decisions on feelings rather than logic and rationality in the context of investing. Further, this heuristic might also explain why media has such an important role in finance. (Slovic et al. 2000, 2-15)
Representativeness Heuristic

Representativeness heuristic is when a person makes a judgment on a sample based on beliefs or patterns from the whole population from where the sample was taken. In other words, the judgment is typically based on traits that belong to a certain stereotype. (Tversky & Kahneman, 1974, 1124)

A example mentioned in Tversky and Kahneman’s essay is the "Steve"-experiment, where Steve is a fictional person described with certain characteristics. In the experiment, the persons were given these details about Steve, a list of possible occupations that could apply for Steve and were finally asked judge what Steve is most likely to work as. The outcome was that the test persons ignored logical arguments such as relative frequency of the possible jobs (i.e. how likely a certain occupation is in general) and only judged the likelihood by the given traits of Steve. As a result, they overestimated the probability of Steve’s work as they focused too much on which stereotype his characteristics belonged to. (Tversky & Kahneman 1974, 1124)

2.1.3 Framing Effect

Framing effect concerns how people’s willingness to taking risk depends on how the situation is presented. The "frame" refers to if something is presented with a positive or negative approach. In general, if the frame is positive, people tend to be less attracted to taking risks. However, when the scenario is portrayed negatively one tends to be more willing to take the riskier alternative. As investing always involves taking on risks, the framing in media should have effect on how people choose to act. (Tversky & Kahneman 1981, 454-455)

2.1.4 Herd Behavior

This heuristic describes the behavior of "following other" instead of following rationality and logic to take a decision. As R.J Schiller mentions in his book "Irrational Exuberance", this behavior is deeply rooted in the animal attributes and the modern society. Schiller claims that buying opportunities in a "hot" stock are likely to spread quickly (Schiller 2000, 151-152). The herd behavior is common in financial markets and is, therefore, liable to have an impact on an IPO. (Bikhchandani & Sharma 2001, 280-281)
3 Mathematical Theory

3.1 Multiple Linear Regression

A linear regression model is specified as

\[ y_i = \sum_{j=0}^{k} x_{ij}\beta_j + e_i, i = 1, \ldots, n \] (1)

Where \( y_i \) is the dependent variable for each observation. Each value of \( y_i \) depends on the covariates \( x_{ij} \) and the error term, also called residual, \( e_i \). The \( x_{ij}:s \) are considered fixed, whilst the \( e_i:s \) are random variables which are assumed to be independent between observations. (Lang 2015, 3)

In our case the dependent variable \( y_i \) is the first-day return and the covariates are hype, stock exchange return, industry, and repo rate. These covariates will be explained and motivated further in the methodology (chapter 4).

The same linear regression can be expressed in matrix notation:

\[ Y = X\beta + e \] (2)

where

\[
Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}, \quad e = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}
\]

and

\[
X = \begin{bmatrix} x_{10} & \ldots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{n0} & \ldots & x_{nk} \end{bmatrix}
\]

3.1.1 Key Assumptions

As described by Hansen (2016, 87-88), the key assumptions for the linear regression model are:

- The observations and covariates are i.i.d.
- The error terms are normally distributed and have a mean of zero i.e. \( \mathbb{E}(e_i|x_i) = 0 \).
• No extreme outliers.
• The error term $e_i$ is independent of $x_i$, i.e. $E(e_i^2|x_i) = \sigma^2$.

3.1.2 Ordinary Least Squares (OLS)

OLS is used to estimate the coefficients, $\beta$, in the regression model. The estimate of $\beta$ is denoted as $\hat{\beta}$, and they are the values that minimizes the sum of the squared residuals $|\hat{e}|^2$, where $\hat{e} = Y - X\hat{\beta}$. $\hat{\beta}$ is found by solving the normal equations:

$$X^T\hat{e} = 0$$  \hspace{1cm} (3)

Inserting $\hat{e} = Y - X\hat{\beta}$ in (3) gives (Hansen 2016, 90)

$$X^T(Y - X\hat{\beta}) = 0$$

$$X^TY - X^TX\hat{\beta} = 0$$

$$X^TX\hat{\beta} = X^TY$$ \hspace{1cm} (4)

$$\hat{\beta} = (X^TX)^{-1}X^TY$$

3.1.3 Dummy Variable

Beside using standard variables as covariates, it is common to use something called a dummy variable. A dummy variable is the same as a binary variable, and takes the value of 1 or 0, depending on a certain criterion. Normally, a dummy variable is employed when testing observations with different qualitative characteristics, for example which industry a company belongs to. (Lang 2015, 20)

3.1.4 Log-transformations in Regressions

In some cases the relationship between the dependent variable $y$ and a covariate $x$ is non-linear. If that is the case, a log-transformation of the covariate makes the relationship non-linear while still being able to perform the regression (Benoit, 2011). The regression model would, in this case, look like:

$$y = \log(x) + e$$

Besides making the relationship non-linear, a log-transformation can also transform skewed data to being approximately more normal.

8
3.2 Model Errors

3.2.1 Heteroskedasticity

As mentioned in section 3.1.1 one of the key assumptions is that the residuals are homoskedastic, which means that they all have the same variance. If the variances are not all the same and depend on one or several of the covariates, it is called heteroskedasticity. An example of heteroskedasticity and homoskedasticity can be illustrated as below:

![Heteroskedastic and homoskedastic residuals](image)

Figure 2: Heteroskedastic and homoskedastic residuals (Wikipedia, 2016)

In the heteroskedastic case, one can see that the residuals’ variances are greater for greater values of the covariate while in the homoskedastic case they appear to be approximately the same.

If the assumption of homoscedasticity fails, the covariance matrix might be
biased. A remedy for this is White’s consistent variance estimator (Hansen 2016, 97-98):

$$\text{Cov}(\hat{\beta}) = (X^T X)^{-1} X^T \left( \sum_{i=1}^{n} x_i^T x_i \hat{e}_i^2 \right) (X^T X)^{-1}$$

(5)

The standard error is the square root of the matrix’ diagonal elements. With this estimate of the covariance matrix OLS can be performed. When the above covariance matrix is used, it is called robust regression.

### 3.2.2 Multicollinearity

Perfect multicollinearity is the case when one or several rows in the matrix $X$ are linearly dependent in some way. More precisely, it means that at least one covariate is linearly dependent of some other covariates. This causes trouble for the OLS method as the linear dependency makes the matrix $X$ non-invertible. Thus the solution $\hat{\beta} = (X^T X)^{-1} X^T Y$ is not possible to derive. (Hansen 2016, 105)

Multicollinearity, not to be confused with *perfect* multicollinearity, is when at least one covariate is significantly correlated with other covariates instead of being completely linear dependent with other covariates. Compared to perfect multicollinearity this is not considered as an error but a disturbance in the model. Multicollinearity results in higher standard deviations of the estimated $\beta$’s of the covariates, and thus making the model more imprecise. (Lang 2015, 15)

An example of multicollinearity is when you run a regression of wage on age, education (number of years) and working experience (number of years). The multicollinearity originates from the fact that $age \approx 6 + education + working\ experience$. (Lang 2015, 15)

The simplest way to solve issues of multicollinearity is to consider which factors that are correlated and find a model where covariates are as uncorrelated with each other as possible. Other remedies are to find new covariates that describe the correlated covariate (these have to be uncorrelated with the original covariates) or to increase the data set. In the case of perfect multicollinearity, the basic solution is to remove the dependent covariates. (Williams, 2015)
3.2.3 Dummy Variable Trap

The dummy variable trap is when the model contains perfect multicollinearity by using all the dummy variables of a certain category in the model. This will cause the intercept to be a linear combination of the dummy variables. There are two ways to solve this, either to delete one of the dummies and use it as a benchmark or to use a model without an intercept. A benchmark can be described as when the model compares the impact of all dummies included in the model with the dummy that was removed (Wooldridge 2009, 227-229). Therefore, the model is benchmarking towards the removed dummy variable.

3.2.4 Endogeneity

Formally, endogeneity is defined as the case when:

$$E(e_i) \neq 0$$  \hspace{1cm} (6)

which is a result of the expected value of $e_i$ being dependent of one or more covariates from the model. The residual is said to be correlated with the covariates. (Lang 2015, 25-26)

In general, four types of common situations generate endogeneity: sample selection bias, simultaneity, missing relevant covariates and measurement errors.

Sample selection bias is when the data has been collected in a certain way which makes the data points biased. In other words, data has not been collected randomly. An example is when a test is conducted to assess if females perform better in school than males. If the tester by some reason selects in particularly talented females, the implications of the study will be misleading as a result of the sample selection bias. (Lang 2015, 26)

One special sort of sample selection bias is called self selection bias. This can be described as when a group by itself is biased even if all the data in that group has been randomly chosen. An illustrative example is when testing if attendance in class has an impact on the final grades. It is probable that the most talented students actively decide not to attend class. If that is the case, then attendance is correlated with talent, and if the talent is part of the residual the attendance is directly correlated with the residual. (Lang 2015, 26)
**Simultaneity** is when the dependent variable has an impact on the covariates. For example, if the dependent variable goes up, this causes the covariate to increase as a result of the increased value of the dependent variable. (Lang 2015, 26-27)

A common example can be found in economics when discussing supply and demand models. For example, imagine that the dependent variable is sold quantities of bananas, and that price represents one covariate. If the amount of sold bananas increases for some other reason than a fluctuation in price (for example an advertisement campaign), then from the theory of economics this causes the price to go up too.

**Missing Relevant Covariates** is self-describing and means that the model is missing a relevant covariate. More precisely, the component of the residual that is correlated with the covariate is lacking in the full model. The remedy is to identify and include the missing component from the residual in the regression model. (Lang 2015, 27)

Finally, measurements errors in the covariates generate endogeneity. It is important to notice that it is only measurement errors in the covariates and not the dependent variable that causes endogeneity. (Lang 2015, 28)

**Remedy for Endogeneity - 2SLS**

The covariate that is correlated with the residual is called endogeneous and a normal approach to handle endogeneity is to create an instrumental variable. This is a new variable that is correlated with the endogeneous variable meanwhile being uncorrelated with the residuals of the model.

The instruments and the non-endogenous variables are called exogeneous variables. To avoid problems of endogeneity, a new matrix is employed, consisting all the exogeneous variables and instruments.

\[
Z = \begin{bmatrix}
  x_{10} & \cdots & x_{1k} & z_{10} & \cdots & z_{1m} \\
  \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
x_{n0} & \cdots & x_{nk} & z_{n0} & \cdots & z_{nm}
\end{bmatrix} \quad (7)
\]

Here, the \(x_{ij}\) consist of all the exogeneous covariates from the model and \(z_{ij}\) to all the relevant instruments. Because of this, \(Z\) can be of same or larger dimensions than \(X\) from 3.1.

In the case where \(Z\) has the same matrix dimensions as \(X\), the normal equation becomes:
\[ Z^t \hat{e} = 0 \]  \hspace{1cm} (8)

However in the other case it is necessary to project \( X \) onto \( Z \) in order to avoid getting an over determined system. This gives:

\[ \hat{X} = Z(Z^t Z)^{-1} Z^t X \]  \hspace{1cm} (9)

where \( \hat{X} \) is the projected part of \( X \) onto \( Z \) and \( X \) is the defined matrix in 3.1. Finally for (9), the normal equation becomes:

\[ \hat{X}^T \hat{e} = 0 \]  \hspace{1cm} (10)

After completing the 2SLS and under the assumption that there is no perfect multicollinearity in \( Z \) or \( \hat{X} \), the regression model is as follows:

\[
Y = Z\hat{\beta} + \hat{\epsilon}, \text{ if number of rows in } Z = \text{ number of rows in } X \\
Y = \hat{X}\hat{\beta} + \hat{\epsilon}, \text{ if number of rows in } Z > \text{ number of rows in } X
\]

The rest follows the standard formulas from chapter 3.1, however with \( X \) replaced with \( Z \) or \( \hat{X} \) in respective cases.

### 3.3 Hypothesis Testing

Hypothesis testing involves tests to verify that a certain result is statistically significant. The interesting part of this study is the point estimates of the linear regression. Formally, for point estimates, the hypothesis test is written as:

\[ H_0 : \theta = \hat{\theta} \]  \hspace{1cm} (11)

where \( \theta \) is the true value of the parameter and \( \hat{\theta} \) its estimate. (Blom et al. 2015, 322)
3.3.1 The F-Statistic

The F-statistic can be used to test the null hypothesis, i.e. that \( \beta_i = 0 \) for some \( i \)'s, and follows a F-distribution under the null hypothesis. Its general definition is given by (Lang 2015, 9-10):

\[
F = \frac{1}{r} \hat{\beta}_r^T V_r^{-1} \hat{\beta}_r
\]  

(12)

In this formulation, \( r \) is the number of tested coefficients. \( \hat{\beta}_r \) represents the vector with the \( \hat{\beta} \) values for the tested coefficients and \( V_r \) is a submatrix of the covariance matrix containing the elements corresponding to \( \beta_r \).

This statistic follows an F-distribution of \( F(r, n - k - 1) \), where \( r \) is, as above, the number of tested \( \beta \)'s. Also, \( n \) is the number of observations, and \( k \) represents the total number of covariates in the model.

Finally, for the one-dimensional case, where only one coefficient is tested, the F-statistic is given by:

\[
F = \frac{\hat{\beta} - \beta_{H_0}}{SE(\hat{\beta})}
\]  

(13)

where \( SE(\hat{\beta}) \) is the standard error of the \( \hat{\beta} \). (Lang 2015, 8-10)

3.3.2 The F-Test

An important assumption for the F-test while testing several coefficients is that the default null hypothesis is that \( \beta \)'s are equal to each other and equal to zero, i.e.

\[
H_0 : \beta_1 = \beta_2 = ... = \beta_r = 0
\]  

(14)

If this is not the case, the usual remedy is to produce a new variable and by using an appropriate variable substitution the model can be re-engineered in such way that the standard F-test assumptions still hold (16), and therefore it is possible to perform the test. (Lang 2015, 10)

After computation of the F-statistic, it is interesting to calculate the probability of getting that F-value for the distribution given that the null hypothesis is true. This is called the p-value and it is computed by:

\[
P\text{-value} = P[F(r, n - k - 1) > F]
\]  

(15)

Where F is the F-statistic. (Lang 2015, 7)
3.3.3 Confidence Interval

By using the one-dimensional F-statistic it is possible to produce a confidence interval, which is given by:

\[ \hat{\beta} \pm \sqrt{F_{1-\alpha}(1, n - k - 1)SE(\hat{\beta})} \]  

(16)

where \( F_{1-\alpha} \) is the \( 1-\alpha \) quantile in the F-distribution and \( \alpha \) is the significance level. (Lang 2015, 7)

3.4 Validation of Model

3.4.1 \( R^2 \) and \( \bar{R}^2 \)

Goodness of fit, also called \( R^2 \), measures how good the data fits a specific model. It gives the relative amount of variation that is explained by the covariates. \( R^2 \) is specified as

\[ R^2 = \frac{|\hat{e}_s|^2 - |\hat{e}|^2}{|\hat{e}_s|^2} \]  

(17)

In this case, \(|\hat{e}|^2\) is the sum of the residuals and \(|\hat{e}_s|^2\) is the residual sum of squares i.e. the sum of squares when \( Y \) is regressed on the intercept only. (Lang 2015, 8)

One problem with the usage of \( R^2 \) is that it increases by the number of covariates in the model. A remedy is to use \( \bar{R}^2 \) which is the same as \( R^2 \) but contains an ad-hoc adjustment for the number of covariates entered in the model. It is defined as:

\[ \bar{R}^2 = 1 - (1 - R^2) \frac{n - 1}{n - p} \]  

(18)

where \( p \) is the number of covariates and \( n \) is the number of observations.

3.4.2 Variance Inflation Factor (VIF)

The variance inflation factor is a statistic to identify and investigate the severity of eventual multicollinearity. Formally, VIF is defined as:

\[ VIF = \frac{1}{1 - R^2_j} \]  

(19)
where $R^2_j$ denotes the $R^2$ for the model in which one of the covariates in the original model is regressed on all the other covariates from the original model. In general, a thumb rule for the VIF-value is that it should not be larger than 10. (Wooldridge 2009, 99)

### 3.4.3 Residual Plot

A residual plot serves to identify certain behaviors of the residuals. Here, it is common to use different plots to see if there is an indication of heteroskedasticity. In general, the residuals are compared to the estimated value of the dependent variable. If the residuals are randomly spread, it would indicate homoskedasticity and if not, the residuals might be heteroskedastic. (Simonoff & Chatterjee, 15-16)

### 3.4.4 Q-Q Plot

The Q-Q plot is used to test the residuals for normality. If the distribution is close to normal, the observations will be plotted close to the straight line in the Q-Q plot. (He, 2015)

### 3.4.5 Leverage and Cook’s Distance

Leverage measures how much a point deviates from the sample mean. As of this, it can be seen as a measure of how much, potentially, an observation might affect the overall result. High leverage for a data point does however not necessarily mean that this data point solely has a large impact on the overall results. (Hansen 2016, 73)

Cook’s distance measures the amount of change in the regression function when one data point is removed. (Jonathan Taylor 2015, 12)
4 Methodology

4.1 Literature

To initiate the study, various literature was studied in order to attain a deeper understanding of the topic. This can be separated into two areas, one denoted as knowledge specific and one denoted as study specific.

The knowledge specific areas concerns learning the relevant knowledge from regression analysis and behavioral finance. The main sources have been Bruce Hansen and Harald Lang with the purpose of attaining an understanding of the most fundamental and relevant regression analysis knowledge.

Regarding understanding the most relevant knowledge in behavioral finance, the book "Thinking Fast and Slow" (2011) by Daniel Kahneman has served as a good source to identify models that are commonly used in the field. Later on, more academic papers on the topics found in Kahneman’s book where searched for.

The study specific part included finding previous academic work where similar investigations have been performed. The main objective of this was to find inspiration on how the project could be designed, especially considering how the subjective variable hype could be quantified.

There were similar studies from where inspiration and ideas have been found. This was important in order to understand what could be expected to investigate, but also to understand the limitations of the study further. One example was that other studies had managed to collect valuable data which was not accessible. For example, the size of discounts in certain cases. Also, this have brought inspiration of how hype can be quantified, for example with Google searches and the trends on Google during the time of the IPO. (Da et al, 2011)

Furthermore, older studies from the field of behavioral finance suggest that the concepts of behavioral finance partially explain the first-day returns of the IPO (Adams et al. 2008). The study by Adams et al. focus mostly on the long-term effect on the stock price, however, the study serves as a good complement. Shefrin (2002) claims that the high first-day returns of IPOs is driven by biases from heuristics. He shows that in "hot" markets, i.e. where other IPOs have been successful, investors will be optimistic and value the IPO above the fundamental value. He also states that, in the long-run, the stock price will fall back and thus will the initial optimism cause a long-term
underperformance of the stock. These papers were useful to find information and inspiration for the case study.

4.2 Data Selection

A list of approximately 250 IPOs completed between 2010 and 2015 was collected from the Nasdaq website. The list contained information about the date of listing, stock ticker, industry, if it is a spin-off, dual-listed or a secondary offering and if the stock is listed on the main market or the growth market First North. The stocks that were spin-offs, dual-listed or secondary offerings were not considered in this study due to the following reasons:

- A spin-off is created by the selling or the distribution of new stocks of an already existing business and by that creating a new independent company. Due to the previous knowledge of the company and stock prices, the bank can set a more accurate price which will have an impact on the first-day return.

- Dual-listed companies was previously listed on another exchange and are therefore not a real IPO.

- A secondary offering spreads the market cap by increasing the number of stocks. This can only be performed after an IPO and hence it is not the first time the stocks are priced or sold.

Google Keyword Tool was used to quantify the hype prior to the IPO. Keyword Tool reports the number of times a certain word was searched for on Google with monthly intervals. The keywords (see Appendix) was chosen on a subjective basis by what an investor is likely to google on when looking for financial information regarding the company. Keyword Tool supplies data from April 2014 to March 2016. Some IPOs from the sample were listed before that and were therefore excluded. All queries was filtered on Sweden, as this is the most relevant region.

The subscription prices were obtained from the Swedish Tax Agency (Skatteverket, 2016) by searching for the company’s name. A few companies did not have their subscription price registered and were therefore not included in the sample.

The first-day closing prices were imported to Google Spreadsheet via the built-in Yahoo Finance API (Market Index, 2016) with the use of each company’s stock ticker.
In the end, around 80 data points were used as some were deleted because of the listing technicalities mentioned above or because of complete data was lacking.

All regressions were performed with the software R.

4.2.1 Removal of Financial Data

One natural aspect to consider during the valuation of the IPO is the financial data. This sort of data was removed from the model by the following motivations:

- Technically, the dependent variable is the difference in the valuation of the stock that the investment bank has set and the valuation that the market has set. The financial data available is the same in both cases. Thus, the difference must be driven by other sources than the financial numbers of the underlying company. As mentioned earlier, this difference is partially explained by the underpricing made by the investment bank and possibly combined with the overall hype.

- Furthermore, the amount of data from financial statements and the amount of data from Google Keyword tools was not completely synchronized. This means that some of the companies had data from Google but it was too old or insufficient. If both Google data and financial data had been used, the number of data points would have decreased to approximately 40.

4.3 Regression Model

4.3.1 Response Variable

First-Day Return

The response variable is the first-day return of the stock, in this study defined as the relative difference between the subscription price and the first-day closing price.

\[
\frac{P_C}{P_S}
\]  \tag{20}

This dependent variable was chosen as it is believed to be the best representation of the short-term hype effects on the stock.
If other trading prices would have been used, for example, the one-week closing price after the IPO, new information about the stock, the overall economy or any other relevant information about the market could have affected the stock's movement. Thus, this would not describe how hype affects the IPO first-day returns.

As of this, the study would rather investigate the effect of hype on the stocks movement rather than the direct effect of hype on the IPO. The overall effect from hype on a stock's movement is also an interesting area to investigate, however for this study it was not part of the scope.

4.3.2 Covariates

**Hype**
Hype has been defined as the relative difference between the number of Google Search Volume, $GSV$, on the company’s keyword the month before the IPO, divided by the monthly average of Google Search Volume, $GSV_{avg}$, before the IPO.

$$\frac{GSV - GSV_{avg}}{GSV_{avg}}$$ (21)

The main assumption is that this difference will represent the hype since the average removes the "standard" interest for this search word. For all search words, the setting has been on searches in Sweden.

The hype is an approximation of the common interest in the stock. However, the covariate does not directly take into account whether the overall opinion towards the stock is positive or negative.

**Nasdaq OMXS30 Monthly Return**
The buy-and-hold return of the OMXS30 index for the month prior to the IPO. This is an indication of how hyped the overall market is during the month leading up to the IPO.

Similar studies suggest that this sort of covariate have had an impact on the first-day return and therefore it was added to the model. (Da et al, 2009)

**Repo Rate**
The repo rate at the IPO. This is an indication of institutional investors'
overall will to invest in the stock exchange, as a high repo tends to lower the interest of stocks and vice versa.

**Industry**
The industry covariates are dummy variables which indicate which industry a company belongs to. As defined by Industry Classification Benchmark (ICB 2016) the industries are:

- Financials
- Consumer Goods
- Consumer Services
- Basic Materials
- Health Care
- Industrials
- Oil & Gas
- Technology
- Telecommunications

The reason for adding these covariates is that some industries might be more hyped than others. For example, the healthcare industry could be considered as more risky since the industry is believed to involve a larger proportion of investments for R&D projects. As of this increased risk, the judgments concerning the companies within the industry might be based more on hopes and expectations than on facts. Therefore, it is reasonable to believe that the hype in some industries affects the first-day returns differently.
4.3.3 Base Model

From the motivations of the covariates and dependent variables, the following base model has been set up to perform the regressions:

First-Day Return = $\alpha + \beta_1$ Hype
+ $\beta_2$ Nasdaq OMXS30 Monthly Returns
+ $\beta_3$ Basic Materials + $\beta_4$ Consumer Goods
+ $\beta_5$ Consumer Services+
+ $\beta_6$ Health Care + $\beta_7$ Industrials
+ $\beta_8$ Oil & Gas + $\beta_9$ Technology
+ $\beta_{10}$ Telecommunications + $\beta_{11}$ Repo

(22)

The regressions are performed in R with the help of White’s consistent variance estimator mentioned in section 3.2.1.
5 Results

5.1 Linear Model

\( R^2 = 0.2212 \)

Table 1: Linear regression

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std.Error</th>
<th>Eta.sq</th>
<th>p.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.041</td>
<td>0.044</td>
<td>0.820</td>
<td>0.000</td>
</tr>
<tr>
<td>Hype</td>
<td>-0.000</td>
<td>0.002</td>
<td>0.000</td>
<td>0.926</td>
</tr>
<tr>
<td>Index returns</td>
<td>1.063</td>
<td>0.524</td>
<td>0.047</td>
<td>0.048</td>
</tr>
<tr>
<td>Basic</td>
<td>-0.096</td>
<td>0.062</td>
<td>0.005</td>
<td>0.129</td>
</tr>
<tr>
<td>Consumer G</td>
<td>0.220</td>
<td>0.134</td>
<td>0.080</td>
<td>0.105</td>
</tr>
<tr>
<td>Consumer S</td>
<td>-0.013</td>
<td>0.082</td>
<td>0.000</td>
<td>0.871</td>
</tr>
<tr>
<td>Health</td>
<td>-0.040</td>
<td>0.083</td>
<td>0.003</td>
<td>0.630</td>
</tr>
<tr>
<td>Industrials</td>
<td>0.045</td>
<td>0.068</td>
<td>0.005</td>
<td>0.511</td>
</tr>
<tr>
<td>Oil</td>
<td>-0.139</td>
<td>0.038</td>
<td>0.006</td>
<td>0.001</td>
</tr>
<tr>
<td>Tech</td>
<td>-0.003</td>
<td>0.093</td>
<td>0.000</td>
<td>0.972</td>
</tr>
<tr>
<td>Repo</td>
<td>-0.227</td>
<td>0.084</td>
<td>0.098</td>
<td>0.009</td>
</tr>
</tbody>
</table>

The coefficient of hype is negative, which is counter-intuitive.

Table 2: Confidence intervals

<table>
<thead>
<tr>
<th></th>
<th>lower</th>
<th>upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.952</td>
<td>1.138</td>
</tr>
<tr>
<td>Hype</td>
<td>-0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Index returns</td>
<td>0.016</td>
<td>2.113</td>
</tr>
<tr>
<td>Basic</td>
<td>-0.220</td>
<td>0.037</td>
</tr>
<tr>
<td>Consumer G</td>
<td>-0.054</td>
<td>0.499</td>
</tr>
<tr>
<td>Consumer S</td>
<td>-0.185</td>
<td>0.155</td>
</tr>
<tr>
<td>Health</td>
<td>-0.213</td>
<td>0.137</td>
</tr>
<tr>
<td>Industrials</td>
<td>-0.097</td>
<td>0.184</td>
</tr>
<tr>
<td>Oil</td>
<td>-0.223</td>
<td>-0.064</td>
</tr>
<tr>
<td>Tech</td>
<td>-0.192</td>
<td>0.188</td>
</tr>
<tr>
<td>Repo</td>
<td>-0.392</td>
<td>-0.061</td>
</tr>
</tbody>
</table>

Table 3: VIF test

<table>
<thead>
<tr>
<th></th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hype</td>
<td>1.19</td>
</tr>
<tr>
<td>Index returns</td>
<td>1.20</td>
</tr>
<tr>
<td>Basic</td>
<td>1.22</td>
</tr>
<tr>
<td>Consumer G</td>
<td>1.46</td>
</tr>
<tr>
<td>Consumer S</td>
<td>1.29</td>
</tr>
<tr>
<td>Health</td>
<td>1.61</td>
</tr>
<tr>
<td>Industrials</td>
<td>1.74</td>
</tr>
<tr>
<td>Tech</td>
<td>1.28</td>
</tr>
<tr>
<td>Oil</td>
<td>1.07</td>
</tr>
<tr>
<td>Repo</td>
<td>1.11</td>
</tr>
</tbody>
</table>

As seen from the VIF-table no covariate had a VIF-value above 10, hence there is no severe multicollinearity.
Considering the negative coefficient of hype, the high p-value and the presence of several points close to Cook’s Distance (in particular observation 63 which had an extremely large hype and negative first-day return) there is a possibility of the correlation between hype and first-day returns being non-linear. Considering these results, a log-transformation of hype was also tested and the result can be seen in the next section.

Figure 3: Plots for linear regression generated in R
5.2 Log-transformation of Hype

\[ R^2 = 0.2278. \] In this regression the \( R^2 \) was slightly larger which is an improvement.

Table 4: Log-transformation of hype

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std.Error</th>
<th>Eta.sq</th>
<th>p.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.021</td>
<td>0.063</td>
<td>0.795</td>
<td>0.000</td>
</tr>
<tr>
<td>log(1+Hype)</td>
<td>0.020</td>
<td>0.038</td>
<td>0.012</td>
<td>0.576</td>
</tr>
<tr>
<td>Index returns</td>
<td>1.181</td>
<td>0.556</td>
<td>0.061</td>
<td>0.042</td>
</tr>
<tr>
<td>Basic</td>
<td>-0.090</td>
<td>0.060</td>
<td>0.000</td>
<td>0.141</td>
</tr>
<tr>
<td>Consumer G</td>
<td>0.221</td>
<td>0.134</td>
<td>0.089</td>
<td>0.100</td>
</tr>
<tr>
<td>Consumer S</td>
<td>0.000</td>
<td>0.084</td>
<td>0.008</td>
<td>1.000</td>
</tr>
<tr>
<td>Health</td>
<td>-0.050</td>
<td>0.095</td>
<td>0.000</td>
<td>0.555</td>
</tr>
<tr>
<td>Industrials</td>
<td>0.050</td>
<td>0.071</td>
<td>0.017</td>
<td>0.486</td>
</tr>
<tr>
<td>Tech</td>
<td>-0.014</td>
<td>0.108</td>
<td>0.000</td>
<td>0.921</td>
</tr>
<tr>
<td>Oil</td>
<td>-0.159</td>
<td>0.044</td>
<td>0.013</td>
<td>0.000</td>
</tr>
<tr>
<td>Repo</td>
<td>-0.210</td>
<td>0.085</td>
<td>0.097</td>
<td>0.010</td>
</tr>
</tbody>
</table>

Table 5: Confidence intervals: Log-hype

<table>
<thead>
<tr>
<th></th>
<th>lower</th>
<th>upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.908</td>
<td>1.132</td>
</tr>
<tr>
<td>log(1+Hype)</td>
<td>-0.041</td>
<td>0.084</td>
</tr>
<tr>
<td>Index returns</td>
<td>0.076</td>
<td>2.297</td>
</tr>
<tr>
<td>Basic</td>
<td>-0.213</td>
<td>0.035</td>
</tr>
<tr>
<td>Consumer G</td>
<td>-0.042</td>
<td>0.483</td>
</tr>
<tr>
<td>Consumer S</td>
<td>-0.164</td>
<td>0.166</td>
</tr>
<tr>
<td>Health</td>
<td>-0.234</td>
<td>0.129</td>
</tr>
<tr>
<td>Industrials</td>
<td>-0.097</td>
<td>0.196</td>
</tr>
<tr>
<td>Tech</td>
<td>-0.215</td>
<td>0.193</td>
</tr>
<tr>
<td>Oil</td>
<td>-0.231</td>
<td>-0.068</td>
</tr>
<tr>
<td>Repo</td>
<td>-0.384</td>
<td>-0.056</td>
</tr>
</tbody>
</table>

Table 6: VIF test: Log-hype

<table>
<thead>
<tr>
<th></th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(1+Hype)</td>
<td>1.18</td>
</tr>
<tr>
<td>Index returns</td>
<td>1.22</td>
</tr>
<tr>
<td>Basic</td>
<td>1.22</td>
</tr>
<tr>
<td>Consumer G</td>
<td>1.47</td>
</tr>
<tr>
<td>Consumer S</td>
<td>1.30</td>
</tr>
<tr>
<td>Health</td>
<td>1.57</td>
</tr>
<tr>
<td>Industrials</td>
<td>1.73</td>
</tr>
<tr>
<td>Tech</td>
<td>1.29</td>
</tr>
<tr>
<td>Oil</td>
<td>1.07</td>
</tr>
<tr>
<td>Repo</td>
<td>1.12</td>
</tr>
</tbody>
</table>

As in the previous chapter, no covariate had a VIF-value above 10. Therefore, any multicollinearity should not be severe.
The Q-Q plot indicates that the residuals are a bit closer to being normally distributed than in the previous regression. Besides that, the effect of the outliers that had a big hype and negative first-day return is reduced considering that the observations are further away from Cook's Distance.
$R^2 = 0.2331$

5.3 Log-transformation of First-Day Returns

The log of the first-day returns was used in an attempt to reduce the heteroskedasticity.

There was no significant difference in the hype’s effect on first-day returns after this transformation.

Table 7: Log-transformation of First-Day Returns

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std.Error</th>
<th>Eta.sq</th>
<th>p.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.028</td>
<td>0.048</td>
<td>0.004</td>
<td>0.570</td>
</tr>
<tr>
<td>log(1+Hype)</td>
<td>0.008</td>
<td>0.026</td>
<td>0.003</td>
<td>0.752</td>
</tr>
<tr>
<td>Index returns</td>
<td>1.069</td>
<td>0.462</td>
<td>0.070</td>
<td>0.025</td>
</tr>
<tr>
<td>Basic</td>
<td>-0.082</td>
<td>0.055</td>
<td>0.006</td>
<td>0.140</td>
</tr>
<tr>
<td>ConsumerG</td>
<td>0.154</td>
<td>0.099</td>
<td>0.060</td>
<td>0.127</td>
</tr>
<tr>
<td>ConsumerS</td>
<td>-0.002</td>
<td>0.075</td>
<td>0.000</td>
<td>0.976</td>
</tr>
<tr>
<td>Health</td>
<td>-0.068</td>
<td>0.077</td>
<td>0.013</td>
<td>0.385</td>
</tr>
<tr>
<td>Industrials</td>
<td>0.026</td>
<td>0.055</td>
<td>0.003</td>
<td>0.642</td>
</tr>
<tr>
<td>Tech</td>
<td>-0.029</td>
<td>0.093</td>
<td>0.002</td>
<td>0.758</td>
</tr>
<tr>
<td>Oil</td>
<td>-0.135</td>
<td>0.036</td>
<td>0.009</td>
<td>0.000</td>
</tr>
<tr>
<td>Repo</td>
<td>-0.181</td>
<td>0.070</td>
<td>0.095</td>
<td>0.012</td>
</tr>
</tbody>
</table>
As can be seen in the Q-Q plot above the observations are closer to the line, and therefore, the log-transformation was partially successful as it reduced heteroskedasticity. However, besides the Q-Q plot, there was no significant difference on hype’s impact compared to the previous regression.
Figure 5: Plots for log-transformation of first-day return generated in R
6 Discussion

6.1 Discussion of Regression Results

6.1.1 Implications From Residual and Q-Q Plot

For the assumption of normally distributed residuals to hold the observations in the Q-Q plot should follow the line. Some of the observations deviate strongly from the line and therefore there is a large possibility of the residuals not being normally distributed.

In the fitted vs residual plot, it is clear that the points follow a certain pattern. The pattern indicates that the data might be heteroskedastic. As of this, the robust regression was used as a remedy.

6.1.2 Implications From Regressions

The results from the three different regressions does not imply that hype has an impact. In all cases, the F-test gave p-values above the limit of 5% for the coefficient of hype. This means that the hypothesis of the true value of the $\beta$’s being equal to zero cannot be rejected. Also, the confidence intervals, at a 95% significance level, of the hype’s $\beta$-value spans from both negative to positive values in both cases, thus it is unclear how hype affects the first-day return.

One covariate that was significant according to this study was the index return. This covariate has a positive impact on the first-day return, which is of interest for individual investors. Even if this was not the primary purpose of this study, the study suggests that an investor should observe how the overall market has moved before the IPO, as this affects investors’ first-day return.

6.1.3 The Linear Model

A positive coefficient was expected to appear in the model as it is believed that a large common interest in a stock would increase the first day return. This was not the case and therefore, in combination with inspecting the data one extra time, a suspicion that one data point was an outlier arised.
However, no direct proofs for this could be found and therefore it was decided not to delete this observation from the sample.

As mentioned in the mathematical theory, Cook’s distance can be used to test if an observation has a large impact on the results. In the graph of fitted values vs. leverage showed that a data point, number 63 (the company was Nuevolution AB), was close the Cook’s distance, and this specific data point might thus impact the attained results. Furthermore, this particular observation had a very high amount of hype compared to other observations and was one of the few observations where the first-day return was negative. Since this point had extreme values compared to the sample’s mean and that there were no logical arguments to delete the data point, the logarithm of hype was taken since this will make potential outliers to have a smaller impact on the results.

6.1.4 Log-transformation of Hype

The results from the log-transformation of hype came out positive. Firstly, the leverage was decreased slightly overall and significantly decreased for data point 63 (Nuevolution AB). Further, the adjusted $R^2$ was improved combined with a decline in p-value by around 40 percentage points. As of this, the log-transformation was considered to be successful.

Finally, the results from the log-modified model were significantly better in comparison to the initial model. However, they were still not good enough to give any clear indications of the true effects from the hype, as the p-value was above the 5% limit as well as most of the confidence intervals contain both negative and positive values.

6.1.5 Log-transformation of First-Day Return

This transformation did not have a major impact on the statistical significance in comparison to the previous regression. However, the residuals were less heteroskedastic.
6.2 Sources of Errors

6.2.1 Mathematical Sources of Errors

As mentioned above, both the Q-Q and residual vs fitted plot did not have optimal results. These results might partly explain why the results are hard to interpret and non-significant.

Moreover, in the final model, around 80 data points were used. It is possible that more data points could increase the study’s significance. This study was limited to the Stockholm markets, however for future research, it could be of interest to scope the study towards larger markets where the IPO frequency is or have been significantly higher.

6.2.2 Complexity of Quantifying Hype

There was one main complication regarding the data with this study, which was finding a good answer on how qualitative attributes such as hype could be quantified in a fair manner.

The most common way to measure hype, based on the previous studies in similar fields, is to use measures such as Google results, Google Trends or Google searches. This is what our study has been based on. However, complications regarding the data have appeared during the project. The Google searches data was presented on a monthly basis. Since it would be logically wrong to embed the interest of stock after the IPO, it was decided that the searches before the month of the IPO should be used. The weakness of this is that the hype the weeks before the IPO was, in most of the cases, not part of the data. Furthermore, this missing part of the hype is expected to be large and therefore, the data is not accurate. If the hype during the same month as the IPO was significantly different for the different companies, this might explain the uncertainty obtained in the results.

This study has not taken into account if the media’s publications have been negative or positive about the stock. This will, as we will mention in the qualitative analysis, probably have an important impact on the first-day return. One reason for that this was not included, was that it would require too much time as well as there existed cases with little media attention.

Due to the Google Searches being filtered on Sweden, the data does not take foreign investors’ hype into account, which may give an underestimate of the
true hype of the company.

6.2.3 Discounts in IPOs

One part that has been mentioned in the theory of an IPO process but not been taken into account to in the mathematical model is that the underwriter normally gives the first investors a discount. The discounts will most likely explain the overall positive first day returns. Since this has not been taken into account to in the model, the assumption is that these discounts are in percentage approximately equally significant. This is in reality not true, however since the specific data of how large the discounts have been from case to case is hard, if not impossible to find.

In further studies, we suggest that collaboration with an investment bank could be beneficial. If the bank has data on all discounts that they have given out, we believe that we could get a better model to see the true impact of hype.

6.2.4 Missing Covariates

One missing aspect that could have brought a less ambiguous interpretation of hype might have been the percentage of the stock’s that was publicly offered. The idea here is that the relative number of stocks represents exclusivity, where a small amount of the total stocks offered would indicate that the stock offering is exclusive. For instance, one can imagine a company that could be assessed in media as the "best investment opportunity in the last decade". If this company would offer a minuscule relative percentage of the stocks to the public market, it is possible that this would increase the hype. This data is usually available when the company makes the IPO. However, it was hard to find years after the IPO.

6.3 Summary

The statistical results fail to confirm that the hype would have an impact on the IPOs’ first-day returns. Several proposed improvements have been suggested for further studies, which might bring stronger conclusions.
Since the study was unable to explain the effect of hype statistically, a case study was performed to try to come to any conclusion regarding the hype’s actual impact on the first-day return.
7 A Case Study of the Media’s Portraying

There are several implications that the stock prices would have increased or decreased in comparison to the subscription price with the support from theories of behavioral finance. An analysis of two cases was conducted and afterwards compared through a discussion.

7.1 The Tobii Case

Tobii is based in Sweden and is a high-tech company within the eye-tracking industry. Tobii was for the time being non-profitable and had had high growth rates before the IPO. The Tobii stock was signed for 25 SEK at subscription and the first-day closing price was 34 SEK, thus a positive first-day return.

In an article published by Privata Affärer, a major financial newspaper in Sweden, it can be concluded that the article is written in a format that is likely to make the reader victims of heuristics. More specifically, these heuristics would increase the hype rather than decreasing it. A few examples are given below: When the article discusses Tobii’s two out of three profitable divisions, it emphasizes that only two out of three divisions are profitable. This can be interpreted as a negative framing of the investment case, and should according to the theory of framing make the investor more likely to taking on risk. This is particularly interesting, as the article finishes with the recommendation to sign into the deal if one likes risky investments. As of this, investors with a riskier profile is by this article probably more positive to the risks than they would have normally been, thus probably also more positive to the investment. (Hemhag, 2015)

Two other heuristics that the reader easily can be exposed to in the same article are the affect heuristic and availability heuristic. The author refers to Tobii as a "growth pearl" which probably generates an association to something valuable or positive. The article also reminds the reader that "almost all IPOs at the Stockholm Stock Exchange has initially increased in value over the last past half year", thus, the risk of overestimating the probability of this to happen is increased (Hemhag, 2015). Therefore, these two examples might make the investor more positive towards the company than he or she normally would have been, and thus more positive to the investment, which drives the first-day return.

Furthermore, SIX Remium anchors the value of 35 SEK per stock without
any logical arguments in an article (Johansson, 2015). This also motivates why investors judgment when valuing Tobii might be higher than it normally would have been, thus making the stock seem like an underpriced stock and driving the hype further.

Finally, one interesting psychological effect that indirectly might have been relevant in this case is the herd behavior. An example was found on the blog Placerapengar where an analysis was published. Here the writer pointed out that serious investors (Investor AB and the Sixth Swedish National Pension Fund) had invested in Tobii. This information might increase the chances of herd behavior as these two players are professional and serious with their investments, which also might explain the final price difference from a behavioral finance perspective. (placerapengar.nu, 2015)

7.2 The Inwido Case

Inwido is a Swedish vendor of windows for commercial as well as private houses in Scandinavia. When Inwido was listed, it had an excellent record of growth (driven by acquisitions) while in the meantime maintaining a stable profit margin. The year before the IPO, the growth was negative, and the sales decreased substantially as a result of strategic changes by Inwido. The Inwido stock was filed for 68 SEK at subscription and the first-day closing price was 64.5 SEK, thus a negative first-day return. (börsdata.se, 2016) (40procent20ar.blogspot.se, 2014)

After reading articles from major financial Swedish newspapers and stock blogs about Inwido it was clear that the analyses were in general based on facts. A few sentences make the reader probable of being exposed to heuristics, however substantially fewer than in the Tobii case.

In Aktiespararna’s analysis, the anchoring of the fair value is below or similar to what Inwido will be valued at, which probably explain why Inwido was valued as a "worse" IPO-signing than Tobii for the individual investor. Furthermore, the analysts does, as in the case of Tobii, take into account that many previous IPOs have been initially traded for higher prices than the subscription price. However, in this analysis, it is emphasized that they in the case of Inwido, do not see any reason or motivation for why this would happen. (Rosenqvist, 2014)

Finally, both companies had a strong history of growth which also was presented differently. In the Tobii case, the word "pearl" was used as a compar-
ison with the growth, which would increase the chances of an investor being a victim of the affect heuristic. This was not the case for Inwido despite that they also had a solid growth history. Some may argue that Inwido started to lose market stocks before the IPO. However, the previous history is still a strong argument of Inwido’s ability to capture growth.

7.3 Summary

By examining and analyzing the main articles just before the IPOs, it is observed, for the case of Tobii and Inwido, that the way information was presented is a possible explanation of how the hype impacts the first-day return.

It is fair to say that Tobii was more hyped than Inwido considering how the stocks were traded relative to the subscription price. This was also reflected in how the both companies were portrayed in Swedish media. Media presented Tobii in a way that made the investor in general more positive to the stock as well as more exposed to heuristics, which in the end could explain why Tobii was more hyped. In the meantime, Inwido was presented differently and more fact based, which made the investor less exposed for heuristics and could make a more rational analysis compared to the Tobii case.

The heuristics presented manages to explain how the hype was built up and how it affected the price differences in the two cases. However, it is also reasonable to believe that the prospect theory played an important role. From the analysis, Inwido was portrayed as if it was a 'bad' investment opportunity meanwhile Tobii was perceived a 'good' opportunity. According to the prospect theory, the 'bad' investment opportunity should make investors more repelled to invest than the 'good' investment would attract investors to invest as they should value the risk of loss more negatively than an equal chance for profit would be valued positively. This might explain why the absolute value of Tobii’s first-day return was relatively larger than Inwido’s.

7.4 Differences in the Two Studies

The both studies give two different implications of the actual impact of hype which is of interest to discuss.
The qualitative study does only concern two cases instead of the 80 that are taken into account in the statistical study. As of this, it is not possible to generalize the qualitative study. As a result of the few number of cases used, the study might have by chance chosen two cases that were successful in terms of explaining how hype affects the first-day returns. In this sense, the study might be somewhat biased. For further research, it would be of interest to expand the qualitative study which possibly could lead to that the both studies would have more similar implications.

Moreover, as mentioned in the mathematical discussion, the statistical study does not take into account if any negative or positive valuations have been included in the media’s portraying. It would be of interest to go through each case’s articles and count negative respectively positive words and use this in the regression analysis. However, this would have required the study to be more extensive than planned.
8 Conclusions

The main aims of this study were to determine to what degree the hype of a company affects the first-day return of a stock on the Nasdaq Stockholm markets. Besides this, the study also aimed to explain how the portraying of a company in the media creates the hype and impact the first-day return.

The main conclusions from the statistical study are vague since the study failed to ensure that hype has an impact at all. The significance of hype was weak, and the results indicate that it could have positive as well as negative or no impact. Further, the study managed to ensure statistically that the overall movement on the market before the IPO has an impact, which is of interest for investors.

The qualitative study cannot be generalized, but the results do still strengthen the hypothesis that media’s portraying and the hype has an impact on the first day returns. This is also aligned with results from the earlier studies mentioned in section 4.1.

To summarize, the study fails statistically to confirm that hype impacts the first-day return. However, the case study manages to strengthen the hypothesis that hype would have an impact, which combined with the discussed errors and limits in this study suggest that further studies should be conducted to validate the impact of hype.
9 References


Börsdata.se. 2016.

Da, Zhi, Engelberg, Joseph & Gao, Penghie. 2009. *In Search of Attention*.


Lang, Harald. 2015. *Elements of Regression Analysis*

Market Index. 2016. *Yahoo Finance API*


Hemhag, Markus. 2015. *Analys: Därför ska du teckna nya IT-bolaget*. Priva-


Rosenqvist, Johan. 2014. *Ingen brådska med att köpa Inwido*. Aktie-
sp经贸. http://www.aktiespararna.se/Artikelarkiv/Borsnotiser/2014/september/Ingen-


40procent20ar.blogspot.se. 2014.
## Appendix

<table>
<thead>
<tr>
<th>IPO firm</th>
<th>Search word</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corline Biomedical AB</td>
<td>Corline Biomedical</td>
</tr>
<tr>
<td>CLX Communications AB</td>
<td>CLX Communications</td>
</tr>
<tr>
<td>Vicore Pharma Holding AB</td>
<td>Vicore Pharma</td>
</tr>
<tr>
<td>Camurus AB</td>
<td>Camurus</td>
</tr>
<tr>
<td>Com Hem Holding AB</td>
<td>Com Hem Holding</td>
</tr>
<tr>
<td>Pegroco Invest</td>
<td>Pegroco</td>
</tr>
<tr>
<td>Byggmästare Anders J Ahlström Holding AB</td>
<td>Byggmästare Anders J Ahlström</td>
</tr>
<tr>
<td>DDM Holding AG</td>
<td>DDM Holding</td>
</tr>
<tr>
<td>Waystream Holding AB</td>
<td>Waystream</td>
</tr>
<tr>
<td>Xbrane Biopharma AB</td>
<td>Xbrane Biopharma</td>
</tr>
<tr>
<td>A City Media AB</td>
<td>A City Media</td>
</tr>
<tr>
<td>Stresscompany AB</td>
<td>Stresscompany</td>
</tr>
<tr>
<td>Nuevolution AB</td>
<td>Nuevolution</td>
</tr>
<tr>
<td>K2A Knaust &amp; Andersson Fastigheter AB</td>
<td>K2A</td>
</tr>
<tr>
<td>Immunovia AB</td>
<td>Immunovia</td>
</tr>
<tr>
<td>Stillfront Group AB</td>
<td>Stillfront</td>
</tr>
<tr>
<td>Hancap AB pref A</td>
<td>Hancap</td>
</tr>
<tr>
<td>VA Automotive i Hässleholm AB</td>
<td>VA Automotive</td>
</tr>
<tr>
<td>Inission AB ser. B</td>
<td>Inission</td>
</tr>
<tr>
<td>Scibase Holding AB</td>
<td>Scibase</td>
</tr>
<tr>
<td>Sprint Bioscience AB</td>
<td>Sprint Bioscience</td>
</tr>
<tr>
<td>Volati AB Pref</td>
<td>Volati</td>
</tr>
<tr>
<td>NP3 Fastigheter AB</td>
<td>NP3</td>
</tr>
<tr>
<td>Nilsson Special Vehicles AB</td>
<td>Nilsson Special Vehicles</td>
</tr>
<tr>
<td>Verisec AB</td>
<td>Verisec</td>
</tr>
<tr>
<td>Advenica AB</td>
<td>Advenica</td>
</tr>
<tr>
<td>GWS Production AB</td>
<td>GWS Production</td>
</tr>
<tr>
<td>Högkullen AB</td>
<td>Högkullen</td>
</tr>
<tr>
<td>Catena Media P,L,C</td>
<td>Catena Media</td>
</tr>
<tr>
<td>SolTech Energy Sweden AB</td>
<td>SolTech</td>
</tr>
<tr>
<td>Gaming Corps AB</td>
<td>Gaming Corps</td>
</tr>
<tr>
<td>TC TECH Sweden AB</td>
<td>TC TECH</td>
</tr>
<tr>
<td>Absolent Group AB</td>
<td>Absolent</td>
</tr>
<tr>
<td>Arcoma AB</td>
<td>Arcoma</td>
</tr>
<tr>
<td>Gränges AB</td>
<td>Gränges</td>
</tr>
<tr>
<td>Company Name</td>
<td>Abbreviation</td>
</tr>
<tr>
<td>------------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Capacent Holding AB</td>
<td>Capacent</td>
</tr>
<tr>
<td>Pandox AB ser. B</td>
<td>Pandox</td>
</tr>
<tr>
<td>Photocat A/S</td>
<td>Photocat</td>
</tr>
<tr>
<td>Hanza Holding AB</td>
<td>Hanza</td>
</tr>
<tr>
<td>Eltel AB</td>
<td>Eltel</td>
</tr>
<tr>
<td>Alimak Group AB</td>
<td>Alimak</td>
</tr>
<tr>
<td>Nilörngruppen AB ser. B</td>
<td>Nilörngruppen</td>
</tr>
<tr>
<td>Tobii AB</td>
<td>Tobii</td>
</tr>
<tr>
<td>Lifco AB ser. B</td>
<td>Lifco</td>
</tr>
<tr>
<td>Magnolia Bostad AB</td>
<td>Magnolia</td>
</tr>
<tr>
<td>Evolution Gaming</td>
<td>Evolution Gaming</td>
</tr>
<tr>
<td>Christian Berner Tech Trade AB</td>
<td>Christian Berner</td>
</tr>
<tr>
<td>Besqab AB</td>
<td>Besqab</td>
</tr>
<tr>
<td>Serendipity</td>
<td>Serendipity</td>
</tr>
<tr>
<td>Kontigo Care AB</td>
<td>Kontigo</td>
</tr>
<tr>
<td>Coor Service Management Holding AB</td>
<td>Coor Service</td>
</tr>
<tr>
<td>Scandi Standard AB</td>
<td>Scandi Standard</td>
</tr>
<tr>
<td>Scandinavian Enviro Systems AB</td>
<td>Scandinavian Enviro Systems</td>
</tr>
<tr>
<td>The Lexington Company AB</td>
<td>Lexington</td>
</tr>
<tr>
<td>Bravida Holding AB</td>
<td>Bravida</td>
</tr>
<tr>
<td>Heimstaden AB Pref</td>
<td>Heimstaden</td>
</tr>
<tr>
<td>LeoVegas AB</td>
<td>LeoVegas</td>
</tr>
<tr>
<td>Collector AB</td>
<td>Collector</td>
</tr>
<tr>
<td>Hoist Finance AB</td>
<td>Hoist Finance</td>
</tr>
<tr>
<td>Inwido AB</td>
<td>Inwido</td>
</tr>
<tr>
<td>Troax Group AB</td>
<td>Troax</td>
</tr>
<tr>
<td>Akelius Pref</td>
<td>Akelius</td>
</tr>
<tr>
<td>Nordax Group AB</td>
<td>Nordax</td>
</tr>
<tr>
<td>Garo AB</td>
<td>Garo</td>
</tr>
<tr>
<td>Maxkompetens Sverige AB</td>
<td>Maxkompetens</td>
</tr>
<tr>
<td>Bactiguard Holding AB</td>
<td>Bactiguard</td>
</tr>
<tr>
<td>Humana AB</td>
<td>Humana</td>
</tr>
<tr>
<td>Attendo AB</td>
<td>Attendo</td>
</tr>
<tr>
<td>Dustin Group</td>
<td>Dustin</td>
</tr>
<tr>
<td>Nobina AB</td>
<td>Nobina</td>
</tr>
<tr>
<td>Capio AB</td>
<td>Capio</td>
</tr>
<tr>
<td>Dometic Group AB</td>
<td>Dometic</td>
</tr>
<tr>
<td>Company Name</td>
<td>Abbreviation</td>
</tr>
<tr>
<td>------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Thule Group AB</td>
<td>Thule</td>
</tr>
<tr>
<td>OrganoClick AB</td>
<td>OrganoClick</td>
</tr>
<tr>
<td>Scandic Hotels Group AB</td>
<td>Scandic</td>
</tr>
</tbody>
</table>