THE ROLE OF VENTURE CAPITAL FUNDING ON IPO UNDERPRICING

A STUDY ON U.S. IPOs BETWEEN 1985 AND 2020

AXEL KOTNIK

JULIUS TÄRNING

Bachelor Thesis

Stockholm School of Economics

2020

The Role of Venture Capital Funding on IPO Underpricing: A Study on U.S. IPOs Between 1985 and 2020

Abstract:

IPO underpricing is a historical phenomenon that has cost entrepreneurs billions of dollars as their companies are undervalued on average when taken public. This paper primarily serves to provide an in-depth analysis on the effect that venture capital-funding has on IPO underpricing when also controlling for VC-firms' bias towards high-tech industries. Secondly, we give a contemporary perspective on the phenomenon by including observations from 2006 to 2020. We show that the framework our methodology is based on exhibits multicollinearity in its original form, and that its unadjusted venture capital-variable likely absorbs the effect of firms being in high-tech industries as a result of the large overlap among their observations. The unadjusted VC-variable is also shown to provide no explanatory power to the model in large. Our constructed non-tech VC-variable indicates, although insignificantly, that venture-capital funding contributes to less underpricing, while the unadjusted VC-variable shows the opposite. This implies that companies and their founders should not assume that venture capital funding damages future IPO prospects by inducing underpricing. Finally, a general increase in IPO underpricing and volatility is observed between 2006 and 2020, while their respective correlations with firm-specific characteristics and own past values have decreased. This suggests that other, unexplored factors have become relatively more important in IPO underpricing in recent years.

Keywords: Initial public offering, IPO underpricing, Venture Capital, Equity Capital Markets, Company Valuation

Authors:
Axel Kotnik (24179)
Julius Tärning (24062)

Tutor:
Adrien d´Avernas, Assistant Professor, Department of Finance

Examiner:
Adrien d´Avernas, Assistant Professor, Department of Finance

Bachelor Thesis
Bachelor Program in Business & Economics
Stockholm School of Economics

Introduction

Initial public offering (IPO) underpricing, which gives rise to a surge in stock price immediately following an IPO, is a phenomenon that has long confounded researchers in finance. Despite its well documented history, there is conflicting evidence on the driving forces behind IPO underpricing, which can be seen not only in its continued occurrence, but in its even larger magnitude today. To date, the most comprehensive and detailed framework for predicting underpricing in IPOs has been constructed by Lowry, Officer & Schwert (2010).

Lowry et al. (2010) has developed a holistic model to predict IPO underpricing, which formulates initial returns as a function of firm-specific characteristics, its previous values in time, and market-wide volatility in asset pricing. Their firm-specific characteristics are proxies for information asymmetry, which by nature should make a company more difficult to value for an underwriter, thus resulting in its offer price being set inaccurately. This inaccuracy most often reflects in *lower* than market valuations in reality since information asymmetry is costly to investors, who are then compensated for this by a surge in stock price the days following the IPO.

One of the examined variables was if a firm is backed by a venture capital-firm (VC) or not. The intuition is that the involvement of a venture capital-firm should mitigate information asymmetry due to them sharing insight and knowledge about the prospect with the underwriter, but previous research on the subject is conflicting; some find that VC-backed firms are more underpriced at listing on average (Megginson & Weiss, 1991) and some find evidence to the contrary (Hasan & Francis, 2001). Lowry et al. (2010) finds, against their hypothesis and in line with Megginson & Weiss (1991), that VC-backing induces underpricing. However, they are skeptical of this result as they believe the variable might unintentionally pick up a “risky industry effect” due to venture capitalists’ bias towards high-risk sectors, whose companies by nature are complex and difficult to value, thus leading to higher underpricing.

Earlier research has thus failed to provide a definite answer as to what the true effect of venture capital-backing on IPO underpricing is, and is worth delving deeper into for the following reasons. Entrepreneurs and their co-investors consistently miss out on up to 20% of equity value due to underpricing, and Ljungqvist (2007) shows that the phenomenon has drained pre-IPO investors of US $125 billion during their 20 years of measurement. For the founders, taking their company public often serves as an exit opportunity, and as such, it is of utmost importance for them to understand the mechanics behind underpricing: that way they can mitigate it and realize maximal returns when going public. While most of the variables measured by Lowry et al. (2010) are out of the entrepreneurs' control to some extent (firm age, industry, underwriter rank, market-wide volatility, etc.), accepting venture capital is a choice, and information about its effect on their IPO prospects is information they can act upon. From a macro-perspective, understanding the effect of VC-backing is becoming increasingly important: between 2005 and 2020, 39% of firms taken public were backed by venture capital in our dataset, a rise from 33% between 1985 and 2005 when including the dot-com bubble and 29% without it (also see appendix A). In summary, the effect of venture capital-backing on underpricing is arguably the most important variable in this model for investors that are eventually looking to go public. It is also playing a increasingly larger role in any model that serves to predict IPO underpricing.

Our contribution is appropriately done following the framework developed by Lowry et al. (2010), as their model is the most sophisticated and comprehensive on the subject as of today.

To capture the true effect of VC-backing, and mitigate the potential "risky industry effect", we alter their VC-variable by disqualifying venture capital-backed IPOs *that also belong to a high-tech industry*. Thus, our proxy for "risky industry" is the tech industry, a decision intuitively motivated by the technological complexity and innovativeness of such firms' products, which limits the predictability of their returns and requires a risk premium for investors. Empirically this is demonstrated in that high-tech firms receive less funding from traditional financial intermediaries (Guiso, 1998) and a larger than average share of venture capital (Lee & Wahal, 2004). This is evident in the overlap of VC-funded and tech-orientation firms in our dataset, in which only 18% of VC-backed firms are not high-tech, and 35% of tech firms are not backed by venture capital. Additionally, we hope to improve the specification of the model and thereby increase the overall understanding of IPO underpricing through this alteration as well: the large intersection in data for the variables Tech and VC raises suspicions of multicollinearity, as Tech is another explanatory variable used by Lowry et al. (2010). Finally, the previously mentioned rise in venture capital funding combined with an increasing share of IPOs being tech companies highlight the increasing importance to correctly specify both Tech and VC. Our research question follows:

> *To what extent does venture capital-funding in companies contribute to higher IPO underpricing and volatility in initial returns when controlled for belonging to a high-tech industry?*

We find that the old VC-variable, which includes tech firms, provides no explanatory power to the model. This implies that entrepreneurs and investors should not assume that VC induces strong IPO underpricing, as implied by the large and positive VC-coefficient in Lowry et al. (2010). However, the true effect of venture capital-funding, i.e. the effect of the non-tech VC-variable, is still uncertain. The large intersection between VC-backed and high-tech firms reduces the number of observations for the non-tech VC-variable to a few hundred, and no significant result is obtained. As a pointer, its value was consistently negative and close to zero, unlike the regular VC-coefficient in Lowry et al. (2010). However, the regular VC-coefficient did converge (although insignificantly) to a negative value close to zero in our final ARMAX(1,1) ARCH(2) regression, thus becoming almost identical to the non-tech VC-variable. The additional data on IPOs from 2005 to 2020 differs from previous years in that it is not as well explained by neither firm-specific characteristics nor its lagged values, despite underpricing and volatility being higher during the period. This suggests that these firm-specific characteristics as proxies for information asymmetry and past average underpricing are less important for predictions today, and even that perhaps other methodologies should be considered for future research. Unlike Lowry et al. (2010), this paper did not investigate overall market volatility as a variable, which may be of larger importance today.

Literature Review

This section will provide an overview of previous findings on IPO underpricing in general, and a summary of the article written by Lowry et al. (2010) in particular, whose theory and research this paper will be based on. Finally, this section will discuss literature on the properties of venture capital funding in IPO contexts.

The fact that IPOs are underpriced on average is well documented in literature and has been studied extensively (Dong, Michel, & Pandes, 2011). Although the underwriter, which sets the IPO offer price, is highly informed about a given IPO prospect's operations and accordingly its equity value, the final price of a firm is always defined by market demand. Therefore, public investors constitute aggregate demand and they will always determine a firm's correct market price, which almost always results in an ex-post price discrepancy (Rock, 1986).

Beatty & Ritter (1986) develops on Rock (1986) by investigating the effect of firm-specific characteristics as uncertainty proxies on IPO offer price. As can be expected, pricing errors are larger in magnitude when companies with uncertainty-yielding characteristics go public. They also find that firms with large pricing errors also tend to be underpriced rather than overpriced, implying that the investors' cost of being uninformed prior to an IPO is later compensated by higher initial returns.

Numerous studies have explored how IPO underpricing has developed over time and the reasons behind the phenomenon itself. Loughran & Ritter (2004) examined trends in IPO underpricing during the period 1990 to 2000 in particular. They found that IPO underpricing increased towards the end, reaching a peak during the dot-com bubble of 1998 to 2000, a period commonly referred to as the "IPO bubble period" in literature.

The article authored by Lowry et al. (2010), published in The Journal of Finance, develops a new and more complete method of evaluating IPO underpricing that incorporates firm-specific asymmetry, time-patterns, and overall market volatility as explanatory variables. Their paper documents IPOs from 1965 to 2005 and examines both IPO monthly *average initial returns* and the *standard deviation of these initial returns*, the latter being an unexplored concept priorly. They find a strong correlation between the two variables across time, i.e. that months with high levels of initial returns also exhibit high volatility in these returns. While the correlation is strongest during the IPO bubble period, they find a significant and prevailing positive correlation across all examined periods. These results demonstrate the complex mechanics behind IPO underpricing and advocates for the use of volatility as part of the equation.

Building on previous literature, they examine firm-specific proxy variables for information asymmetry (uncertainty) as explanatory variables for the level and volatility of IPO underpricing. While they considered most variables to be relatively uncontroversial in terms of their definitions and results, they became doubtful of their VC-variable. They suspected that it mistakenly picked up a "risky industry effect" since venture capital firms are often biased towards industries defined by high growth and riskiness. This proposition was not examined further.

The empirical evidence of the effect that venture capital funding has on IPO underpricing is conflicting. Megginson & Weiss (1991) find that firms backed by venture capital experience significantly lower underpricing than those who receive no venture capital funding. Evidence presented by Hasan & Francis (2001) suggests the complete opposite, i.e. that firms who receive funding from venture capital firms tend to have higher initial returns than the control group on average.

Contribution and Limitations

The results presented by Lowry et al. (2010) suggest that IPO underpricing is related to both firm-specific factors and market-wide factors. Our intention is to contribute to this article and prior research in two ways:

(1) We will examine a larger time set, by extending the data to 2020. In this regard, our paper is intended for anybody interested in the behavior of IPO underpricing and volatility from a more recent perspective. We will combine these new results with a methodological replication on the same time set of Lowry et al. (2010), to ensure comparability.

(2) To isolate the true effect of venture capital on IPO underpricing, i.e. eliminate the risky industry effect, we will replicate the regressions in Lowry et al. (2010) with a VC variable that excludes high-tech firms. By defining "risky industry" as high-tech industries, we also hope to improve the model in its entirety by eliminating potential multicollinearity.

For our thesis, we will not reconstruct the section in Lowry et al. (2010) that incorporates market-wide uncertainty in equity securities as an explanatory variable for IPO underpricing. Although this limits the scope of our paper, it is not relevant with regards to the effect of venture capital-backing in IPOs.

Another limitation is that we lack information regarding firms' age and underwriter ranks, both of which were firm-specific proxies for information asymmetry in Lowry et al. (2010). As an implication, there may be slight discrepancies between the coefficients in our regressions and those in Lowry et al. (2010).

Methodology

We collected our data from the Securities Data Company database (SDC Platinum), which contains information on IPOs ranging from 1985 to 2020 on the U.S. stock market. Lowry et al. (2010) mainly used SDC data in their article as well, but complemented this with a few hand-collected datasets. We chose not to include these as they (1) mostly covered IPOs prior to 1985, (2) only a negligible share of their data came from these sources, and (3) because the datasets often covered the same IPOs. As can be seen in Table I, the unfiltered file from SDC provides us with 10 058 observations. 7 820 of the observations are from the period 1985 to 2005 and 2 238 observations are from the period 2006 to 2020.

In total, we have removed 2 745 misguiding observations. This includes all class A-issues, for which there are always corresponding class B-issues, which otherwise would result in the double-counting of some IPOs. In addition to this, we have removed shares that were traded over-the-counter (OTC), as well as different types of financial funds and instruments found in the original dataset. Lastly, we have removed a handful of observations containing faulty information, including stocks being listed on a non-American exchange or those with apparent incorrect information about their offer price or stock price.

With these adjustments in place, we arrived at a total of 7 312 unique IPOs with 5 785 observations from 1985 to 2005 and 1 527 observations from 2006 to 2020.

Table I
Sources of IPO data, 1985-2020

Below is a summary of our dataset from the different subperiods. Data available is defined as class-B issues on Nasdaq or NYSE with correct data.

Data Source	Sample Period	Number of IPOs	Data Available
Securities Data Company	1985-2005	7820	5785
Securities Data Company	2006-2020	2238	1527
Securities Data Company	1985-2020	10058	7312

The specific information we have collected for the firms taken public are: (1) the final price of the issue; (2) the stock price four weeks after the first trading day; (3) if the firm is listed on Nasdaq or the New York Stock Exchange; (4) if the company's core business is defined as high-tech (for instance, if a business model focusing on telecommunication or data processing services); (5) if the firm has received venture capital; (7) the total number of shares issued, and; (8) the middle range offer price as filed in the company's IPO prospectus.

After filtering the dataset, we have grouped the IPOs by the month they were taken public. For example, if ten companies performed their IPO during March 2007, their average value is used as a single data point defined as "March 2007". Although our data points are ordered chronologically in our analysis, the fact that some months had no IPOs at all means our data is not always equally spaced in time.

This paper has been structured into three separate chapters whose results are contingent upon each other in the following order:

I. Descriptive Analysis

II. Cross-Sectional Analysis
III. Time-Series Analysis

I. Descriptive Analysis

Following the methodology of Lowry et al. (2010), we define an IPO's initial return (IR) as the percentage difference between the offer price and the stock price four weeks after listing to avoid the influence of underwriters' price stabilization measures on the stock price the immediate days following the offering (Ruud 1993; Hanley, Kumar, & Seguin 1993). Our data supports this decision as well; using the stock price one day after the offer, we find that 57% of IPOs have initial returns between -10% to 10%, compared to only 43% if we use the stock price one month after. Furthermore, the share of IPOs with initial returns less than -10% is substantially smaller one day later compared to one month later (4% versus 13% respectively). Therefore, we use the stock price 31 days after going public.

For a given month, i.e. a given data point, the term "average IPO initial return" is defined as follows:

$$\overline{IPO\ initial\,return_i} = \frac{1}{n}\sum_{ij=1}^{in}\frac{stock\,price_{ij} - offer\ price_{ij}}{offer\ price_{ij}}$$

where i denotes month i, n denotes the total number of companies going public that month, and j denotes the IPO of company j

By grouping our data into monthly averages, we can test for correlations between average IPO initial returns and the standard deviation of IPO initial returns for a single data point. The monthly standard deviation of IPO initial returns is defined as the standard deviation of that month's individual companies' IPO initial returns. Without grouping our data into months, we would only have a single initial return per data point, which cannot yield a standard deviation.

For a given month, i.e. a given data point, the term "standard deviation of IPO initial returns" is defined as follows:

$$IPO\ initial\ return\ \sigma_i = \sqrt{\frac{\sum_{ij=1}^{in}(IPO\ initial\ return_{ij} - \overline{IPO\ initial\ return_i})}{n-1}}$$

where i denotes month i, n denotes the total number of companies going public that month, and j denotes the IPO of company j

Throughout our sample, twelve separate months contain only one IPO, which means there are no monthly standard deviations for these data points. This is the tradeoff between using larger groups (e.g. quarters, half-years, years), which generates standard deviations for every period, and using smaller groups (e.g. months), which generates more data points. Our decision to group monthly partially stems from following the methodology of Lowry et al. (2010), and partially from our desire to generate more insightful findings from having more data points.

Lastly in this section, we have performed autocorrelation tests with different lags for the two variables average and standard deviation of monthly IPO initial returns. This will provide an indication as to whether the two are correlated and how the data behaves over time.

II. Cross-Sectional Analysis

In this section, we examine firm-specific characteristics that may explain the results from section I. The variables were chosen following Lowry et al. (2010) and act as proxies for information asymmetry in a company in anticipation of its IPO. More information asymmetry should increase the returns and standard deviation of returns after an IPO due to the cost of being uninformed as an investor. For our extension, we have added a non-tech VC-variable.

Four of our explanatory variables in our raw data set are dummy variables (VC, non-tech VC, Tech, NYSE). For example, a company is either backed by venture capital (VC = 1) or it is not (VC = 0). To transform these into monthly, non-binary variables that are useful in regressions, we calculate each month's percentage share of firms with a given quality. For example, in a month where three out of four companies are backed by venture capital, our VC-variable takes a value of 0,75.

Two of our explanatory variables are continuous variables for each individual company (Log(shares), Absolute price update). To transform these into monthly variables, the average variable value amongst the firms in the month is used.

Firstly, we examine the variables' individual correlations with the two dependent variables (the mean and standard deviation of IPO initial returns) over time. This provides an indicative estimate of their significance and explanatory value. Secondly, we model the variables jointly through regressions on the mean and standard deviation of IPO initial returns. This is done through both ordinary least-squares (OLS) regressions and maximum likelihood-estimations (MLE) of coefficients generated in weighted least-squares (WLS) regressions using the standard deviations of the residuals as weights. For the WLS, the log of the variance of the residuals is assumed to be linearly related to the firm-specific characteristics, which circumvents the problem of receiving biased standard errors if the variance of the residuals is not constant. The fit of the two models will be compared using log-likelihoods and Akaike Information Criterions (AICs), which is essentially log-likelihoods penalized for added model complexity (i.e. more explanatory variables).

The equation for the mean of IPO initial returns in the WLS model follows:

$$IR_i = \beta_0 + \beta_1 Log(shares_i) + \beta_2 Tech_i + \beta_3 VC_i + \beta_4 NYSE_i + \beta_5 |Price\ Update_i| + \varepsilon_i$$

$$IR_i = \beta_0 + \beta_1 Log(shares_i) + \beta_2 Tech_i + \beta_3 non\text{-}tech\ VC_i + \beta_4 NYSE_i + \beta_5 |Price\ Update_i| + \varepsilon_i$$

The equation for the natural logarithm of the variance of the residuals in the WLS model follows:

$$Log(\sigma^2(\varepsilon_i)) = \gamma_0 + \gamma_1 Log(shares_i) + \gamma_2 Tech_i + \gamma_3 VC_i + \gamma_4 NYSE_i + \gamma_5 |Price\ Update_i|$$

III. Time-Series Analysis

In the final section of our analysis, we model our cross-sectional data as a time-series process. This is based on the theory that in times when there is a high degree of uncertainty and mispricing in IPOs (i.e. high average and standard deviation of IPO initial returns), concurrent valuations are less reliable and thus the following months should be characterized by uncertainty as well (Lowry et al., 2010). In short, the average and standard deviation of IPO initial returns are now modeled as a function of both firm specific characteristics *and* their own lagged values.

Potential time-series patterns in the mean of IPO initial returns will be modeled with an autoregressive moving average model ARMAX(*p,q*), which is a Box & Jenkins (1976) ARMA-process that includes explanatory variables in the regression. The autoregressive term AR(*p*) models the structural disturbances (as the model assumes some parts of the residuals to be predictable, these are called structural disturbances) as a process of their lagged values. The moving average term MA(*p*) models the white noise-disturbances in the residuals as a process of past values. Consider a regression with time-dependent residuals μ_t:

$$y_t = \beta_0 + \beta X_t + \mu_t$$

An ARMAX(*p,q*)-process assumes the residuals μ_t follow:

$$\mu_t = \underbrace{\left(\phi_1\mu_{t-1} + \cdots + \phi_p u_{t-p}\right)}^{\text{AR-terms}} + \overbrace{\left(\epsilon_t - \theta_1\epsilon_{t-1} - \cdots - \theta_q\epsilon_{t-q}\right)}^{\text{MA-terms}}$$

Where ϵ_t are the residuals of the residuals, generated by a random white-noise process.

To specify the ARMAX-model correctly, it is required to make sure the series is stationary (i.e. the level of initial returns remains roughly constant throughout our data). As can be seen in Figure 2 further into the paper, the initial returns in our sample are sufficiently stationary. The order of *p* (how many lags the AR-term should include) and the order of *q* (how many lags the MA-term should include) is set as the last significant lag (with 95% confidence) in an autocorrelation function and partial autocorrelation function of the data, respectively. For our data, both *p* and *q* are determined to be optimally set as 1.

Potential time-series patterns in the standard deviation of IPO initial returns will be modeled with an autoregressive conditional heteroskedasticity ARCH(*m*) model (Engle, 1982), which models the variance of the residuals, rather than the residuals themselves, as a Box & Jenkins (1976) autoregressive (AR) process. In a regular OLS regression, the conditional variance of the error terms is assumed to be constant and equal to the unconditional variance. With an ARCH(*m*)-process, the conditional variance of the error term is not constant and depends on prior error terms, but the unconditional variance is still constant. Consider a regression with time-dependent residuals μ_t:

$$y_t = \beta_0 + \beta X_t + \mu_t$$

An ARCH(m)-process assumes the variance of the residuals follows:

$$Var(\mu_t) = \sigma_t^2 = \omega + (\alpha_1 \mu_{t-1}^2 + \cdots + \alpha_m \mu_{t-m}^2)$$

In addition to the using AIC-scores and log-likelihood to measure the fit of our models, Ljung-Box tests (Ljung & Box, 1978) will be performed on the autocorrelation function of the models' predictions. Ljung-Box tests measure the autocorrelation of the residuals and is thus a diagnosis for how well a model functions with regards to its own, past values. If autocorrelation has been modelled adequately, the Ljung-Box should become lower.

Empirical Results

I. Descriptive Analysis

Figure 1 shows the summary statistics for the individual companies' IPO initial returns in our sample, in addition to a normal distribution with identical mean and standard deviation. The data is defined by high kurtosis and a positive skew, and 64% of observations are found to have an initial return between -10% and 30%.

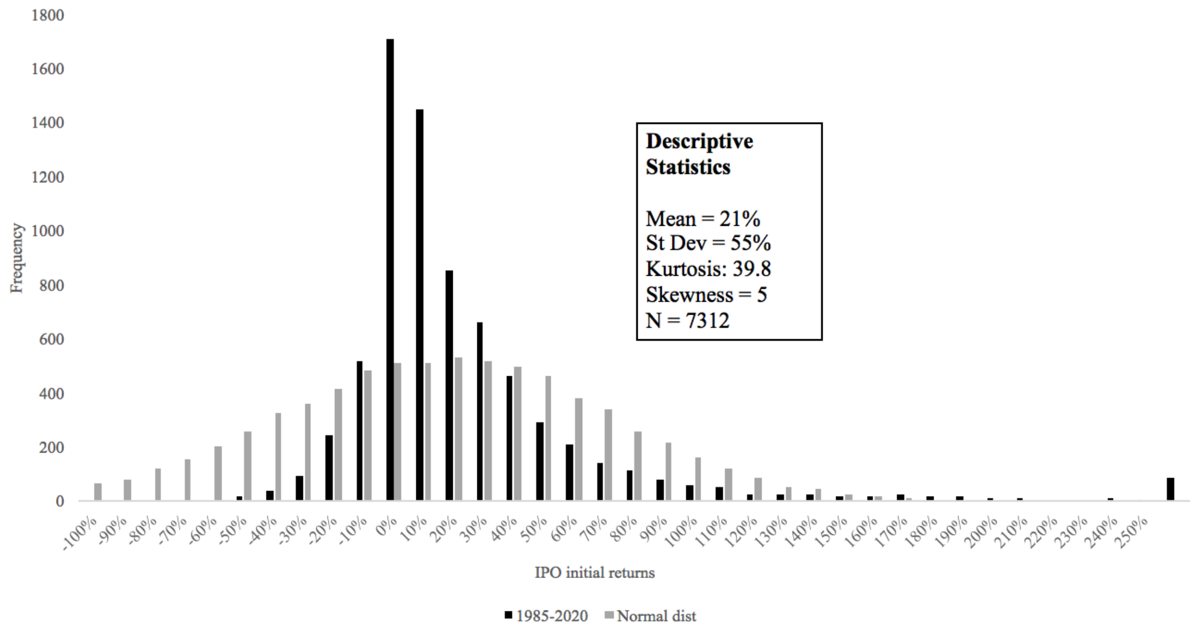

Figure 1. ***Frequency distribution of IPO initial returns (1985 -2020).*** Histogram displaying the distribution of IPO initial returns for our sample compared with a normal distribution.

Figure 2 displays monthly means and standard deviations of IPO initial returns from 1985 to 2020. Visually the two appear to be strongly positively correlated, where a month with high initial returns on average also has a high dispersion of returns, which is in line with Lowry et al. (2010). This is confirmed by testing their correlation, as seen in Table II. For our entire sample period, the correlation between average IPO initial returns and the cross-sectional standard deviations of these returns is 0,886. Looking at Figure 2, the time period 1998 to 2000 is characterized by uniquely high average initial returns and standard deviations. This will, in line with Loughran & Ritter (2004), hereafter be referred to as the "IPO Bubble Period" (September 1998 to August 2000). By including this period, the correlation between the level and volatility of IPO initial returns dramatically rises, which is why it is excluded or captured in a separate dummy variable in subsequent analyses. For our time extension, there is no need to classify another bubble period, as the level and volatility in initial returns seem to be fairly modest and consistent. We do find a somewhat general increase in mean and standard deviation across the extended time set however.

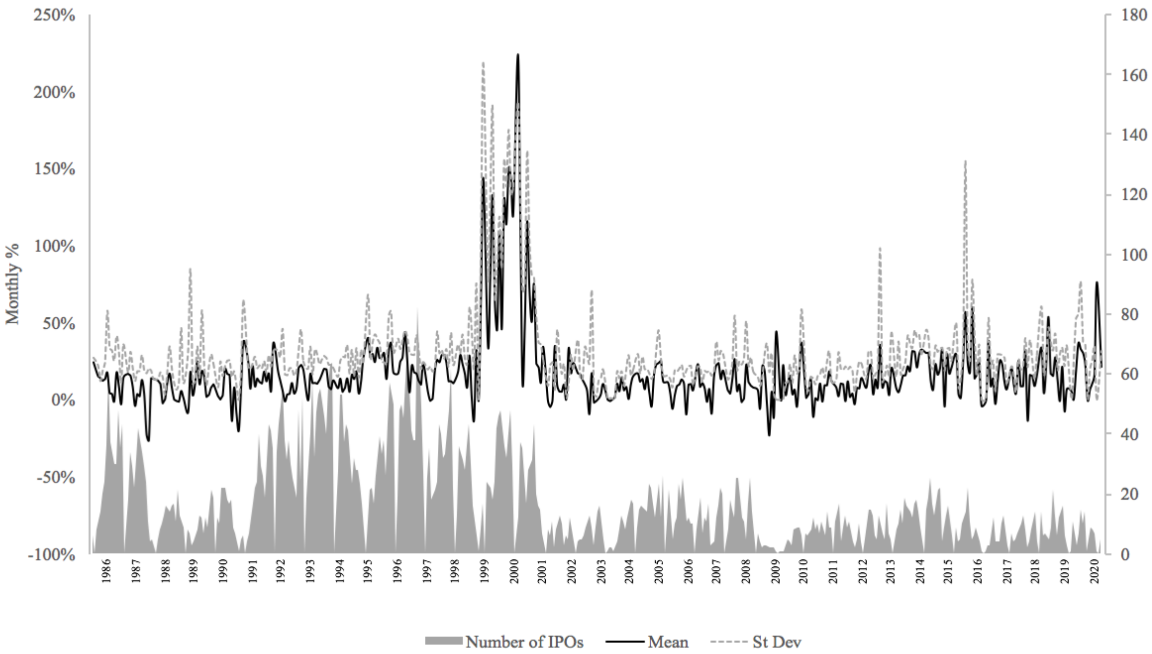

Figure 2: ***Monthly mean and standard deviation of IPO initial returns and the number of IPOs by month.*** The mean for a given month is calculated as the average initial return for a month, defined as the average percentage change between the offer price and the stock price 31 days later for all companies that went public that month. The standard deviation for a given month is calculated on the initial returns of the companies in that month.

Table II provides further insight into the characteristics of the monthly means and standard deviations of initial returns that are displayed in Figure 2. As can be seen in Table II, we have divided the dataset in four different time periods.

We find the correlation between IPO initial returns and the standard deviation of IPO initial returns to be strong across the four examined periods. The correlation for our time extension (2006 to 2020) is 0,773. When assessing the entire time sample (1985 to 2020) we find a correlation of 0,886. The correlation is slightly weaker, 0,684, when excluding the IPO bubble period (August 1998 to September 2000). These results are consistent with the visual representation seen in Figure 2.

To gain an insight into our extended time period 2006 to 2020, it is most useful to compare it to the entire sample period while omitting the bubble period. Between these two subsets, the average and mean of IPO initial returns has increased in 2006 to 2020.

Table II

Descriptive Statistics on the monthly mean and volatility of IPO initial returns (with lags)

Descriptive statistics on the cross-sectional averages and standard deviations of monthly IPO initial returns are displayed. The correlations (Pearson correlation coefficient) are calculated on the average and the standard deviation (σ) of monthly initial returns.

	N	Mean	Median	Std. Dev.	Corr.	Autocorrelations: Lags 1	2	3	4	5	6
1985-2020											
Average IPO initial return	392	0,176	0,123	0,261		0,62	0,46	0,58	0,55	0,47	0,48
Cross-sectional σ of IPO IR	405	0,320	0,244	0,293	0,886	0,63	0,55	0,61	0,63	0,56	0,53
1985-2005											
Average IPO initial return	237	0,196	0,124	0,302		0,72	0,55	0,66	0,64	0,54	0,53
Cross-sectional σ of IPO IR	241	0,346	0,244	0,340	0,910	0,74	0,67	0,70	0,69	0,62	0,62
1985-2020 omitting bubble											
Average IPO initial return	369	0,130	0,116	0,145		0,26	0,26	0,11	0,04	0,10	0,14
Cross-sectional σ of IPO IR	379	0,264	0,230	0,161	0,684	0,15	0,05	0,13	0,16	0,15	0,11
2006-2020											
Average IPO initial return	155	0,145	0,119	0,179		0,12	0,06	0,10	-0,02	0,09	0,13
Cross-sectional σ of IPO IR	162	0,280	0,244	0,197	0,773	0,14	0,02	0,12	0,18	0,13	0,05

Both variables exhibit strong autocorrelation when assessing the full time period and including the bubble period (1985-2020). When excluding the bubble period (1985-2020 omitting bubble), we see a significant drop in autocorrelation amongst average IPO initial returns and cross-sectional standard deviation respectively, which suggests bubble periods could be classified as periods with (1) high means and high standard deviations of IPO initial returns, and (2) high autocorrelation. The autocorrelations in the extended time period 2006 to 2020 is lower than in all other time periods for all lags, and close to non-existent for some lags. Although the time period 2006 to 2020 displays high returns and standard deviations, it also exhibits very low autocorrelation and will thus not be considered another bubble period. The time-dependent factor of IPO returns will be explored further in section III.

II. Cross-Sectional Analysis

Previous literature (Rock, 1984) has proposed the idea that some companies are characterized by larger information asymmetry prior to their IPOs, and should as such experience higher variability in their immediate returns. Additionally, ex-ante information asymmetry should yield higher initial returns as investors are compensated for the cost of being uninformed.

What observable factors impact the information available about a given firm prior to its IPO? An established firm that has been covered by analysts for years should theoretically be easier to price for an investment bank a priori to its IPO. In contrast, a young, high-technology firm whose valuation mostly depends on its future growth potential should theoretically be more difficult to price. Suppose some periods on average are characterized by a higher share of IPOs with characteristics that in theory makes them more difficult to value. This would, as shown by Lowry et al. (2010), result in some periods having higher initial returns and higher initial return volatility.

In order to quantify this reasoning and make it measurable, we will examine the relation between initial returns and a set of variables that each should contribute to making an IPO more difficult to value. Following Lowry et al. (2010) to the best of our ability, given the constraints of our dataset, we have constructed a list of *firm-specific characteristics* in Table III.

Table III

Definition of variables used in our regression

Displays what values the variables take for each firm depending on its firm-specific characteristics.

Variable	Definition
NYSE dummy	Is equal to 1 if listed of New York Stock Exchange and 0 if listed on NASDAQ
Tech dummy	Is equal to 1 if the firm in a high-tech industry and 0 otherwise
VC dummy	Is equal to 1 if the firm is backed by a venture capital firm and 0 otherwise
Non-tech VC	Is equal to 1 if the firm is a no-tech firm and backed by a venture capital firm and 0 if it is a tech firm that recievs funding, also 0 if there is no venture capital firm involved
Log(Shares)	Taken the logarithmic value of shares offered (in millions)
Average Price Update	The absolute value of the percentage change between the middle of the range of prices in the initial offering and final offer price
Bubble period dummy	Is equal to 1 if between September 1998 and August 2000

The theoretical hypotheses as to why these characteristics should have an effect on an underwriter's ability to accurately value a given firm follow (Lowry et al., 2010):

NYSE. The New York Stock Exchange (NYSE) is considered more prestigious and comes with a heftier price tag. This results in more established firms being listed on NYSE and more growth-oriented firms being listed on other exchanges, which should impair the underwriter's ability to correctly value firms listing on other exchanges. Our hypothesis is thus that the average and standard deviation of initial returns should be lower for firms listed on NYSE.

Tech. Valuing tech firms are often more difficult due to the uncertainty of their future growth prospects (Kohers & Kohers, 2004). A general misunderstanding of technological innovations and their predicted success is the main factor that drives uncertainty (Luigi, 1998). Therefore, our hypothesis is that tech-oriented firms are more difficult to value, and tech firms will have higher and more volatile initial returns on average.

VC. Lowery et al. (2010) made the initial assumption that firms should be easier to value when venture capital firms are involved due to their extensive due diligence on prospective investments. However, following their opposite results, and due to the potential risky industry effect and large overlap amongst tech and VC-backed firms in our sample, we hypothesize that this variable will follow the Tech variable closely both in terms of magnitude and direction, and thus the VC variable should increase the average and standard deviation of IPO initial returns.

Non-tech VC. The non-tech VC variable will attempt to diminish the risky industry effect that the former VC variable may suffer from and thus capture the true nature of VC firms' effect on IPO initial returns. Our assessment, as detailed previously, is that the tech industry is the best proxy for "risky industry"[1]. Our hypothesis is that non-tech VC as a variable diminishes information asymmetry and thus yields lower averages and standard deviations of IPO initial returns.

[1] One could argue that the reverse method could be used, i.e. having a non-VC tech-variable. However, this is not the case as the business idea (if it is tech or not) comes before the company receives venture capital-backing.

Log(shares). Underwriters should theoretically experience difficulty in valuing smaller offerings since smaller companies naturally cannot offer as detailed information a priori to their IPO as larger ones. Thus, our hypothesis is that months with larger share offers should experience smaller and less volatile initial returns.

Absolute Price Update. This parameter can be regarded as a proxy for how much new information is learned about the firm during the registration period. Higher uncertainty about a given company's value should therefore translate into a higher value for this variable, and thus result in higher averages and standard deviations of initial returns.

In Table IV we outline the correlations between average initial returns and standard deviation of initial returns with the firm-specific characteristics. Firstly, one observes that most of the firm-specific characteristics are not significantly correlated during the 2006 to 2020 period. Accordingly, we will not analyze this period by itself from now on, but solely as a part of our entire sample from 1985 to 2020. The total period will be compared to the earlier subperiod 1985 to 2005 to gain insight into the changes that have occurred from 2006 and onwards, and also for comparison with the results from Lowry et al. (2010), whose time period is 1981 to 2005.

Table IV

Correlations between monthly average initial returns and standard deviations and monthly market characteristics

The Pearson correlation coefficients are calculated for the average and standard deviation of monthly initial returns and the average monthly characteristics. For the dummy variables (NYSE, Tech, VC, non-tech VC, NASDAQ), the monthly characteristic is calculated as the percent of IPOs in that month that were categorized as 1. For the continuous variables (Log(shares), Price update), the monthly characteristic is calculated as the average value of the companies' values for those variables. The parentheses below the correlation coefficients are robust p-values.

Specifications	1985 - 2005				2006 - 2020	
			Omitting bubble			
	Average IR	StDev IR	Average IR	StDev IR	Average IR	StDev IR
Percent NYSE	-0,43	-0,43	-0,41	-0,39	-0,12	-0,21
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,2558)	(0,0057)
Percent Tech	0,61	0,59	0,44	0,39	0,29	0,16
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0028)	(0,0318)
Percent VC	0,57	0,56	0,30	0,29	0,34	0,19
	(0,0000)	(0,0000)	(0,0000)	(0,0001)	(0,0012)	(0,0172)
Percent non-tech VC	-0,07	-0,08	0,08	0,01	0,15	1,00
	(0,2756)	(0,1972)	(0,2726)	(0,9343)	(0,9052)	(0,9590)
Average Log(shares)	0,15	0,18	0,16	0,24	0,07	-0,05
	(0,0010)	(0,0000)	(0,0055)	(0,0002)	(0,4205)	(0,4168)
Average Price Update	0,42	0,39	0,26	0,21	0,04	-0,07
	(0,0003)	(0,0001)	(0,0004)	(0,0112)	(0,5882)	(0,3868)
Percent NASDAQ	0,43	0,43	0,41	0,39	0,12	0,21
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,2558)	(0,0057)

Specifications	1985 - 2020			
			Omitting bubble	
	Average IR	StDev IR	Average IR	StDev IR
Percent NYSE	-0,33	-0,36	-0,25	-0,29
	(0,0000)	(0,0000)	(0,0002)	(0,0000)
Percent Tech	0,47	0,44	0,37	0,27
	(0,0000)	(0,0000)	(0,0000)	(0,0000)
Percent VC	0,44	0,40	0,34	0,25
	(0,0000)	(0,0000)	(0,0000)	(0,0000)
Percent non-tech VC	-0,04	-0,05	0,03	0,01
	(0,2876)	(0,1788)	(0,5541)	(0,8641)
Average Log(shares)	0,13	0,13	0,10	0,07
	(0,0004)	(0,0007)	(0,0533)	(0,2141)
Average Price Update	0,36	0,31	0,16	0,07
	(0,0005)	(0,0005)	(0,0033)	(0,2135)
Percent NASDAQ	0,33	0,37	0,25	0,29
	(0,0000)	(0,0000)	(0,0002)	(0,0000)

A quick overview of Table IV tells us that all correlations are weaker in our entire period, 1985 to 2020, when compared to the 1985 to 2005 subperiod. Additionally, no particular variable has decreased or increased in relative magnitude either. This indicates that other factors beyond information asymmetry in individual IPOs may have become more important in later years. It appears not to be an increased time-dependency either as the autocorrelations in Table II were lower when compared to all other periods, although this will be further explored in section III.

In line with our hypothesis, by looking at Percent Tech and Percent VC together and Percent non-tech VC separately, we see suggestive evidence of multicollinearity between the two first variables. This can be seen in the similar magnitude of the variables in all columns (e.g. 0,57 and 0,61 in 1985 to 2005; 0,29 and 0,34 in 2006 to 2020; 0,47 and 0,44 in 1985 to 2020 respectively). Indeed, the bivariate correlation between the two variables is 0,78, thus confirming multicollinearity to some extent. Regarding the non-tech VC-variable, its coefficient is close to zero for most time periods, as opposed to the VC-variable which takes on values between 0,25 and 0,57. The limited amount of observations that are both non-tech and venture capital-backed (457 in total) hinders us from drawing strong conclusions on its true value, which can also be seen in its insignificant p-values. However, we will include it in our coming regressions to gain a perception of its value and see if it is modeled better in other specifications.

Finally, a noteworthy but expected observation in Table IV is also that the omittance of the bubble period weakens the coefficients for almost all variables. For our continued analysis, the bubble period will be captured by a bubble dummy variable instead, as it allows us to analyze the period while simultaneously keeping it from interfering with the other variables.

To model the firm specific characteristics jointly, we have performed OLS and WLS regressions in Table V. The OLS regression will serve as a baseline model for comparison against the heteroskedasticity-consistent WLS regression. The WLS will likely be a better fit for the data, as can be seen preliminarily in the heteroskedasticity amongst the residuals in the regular OLS model in Figure 3.

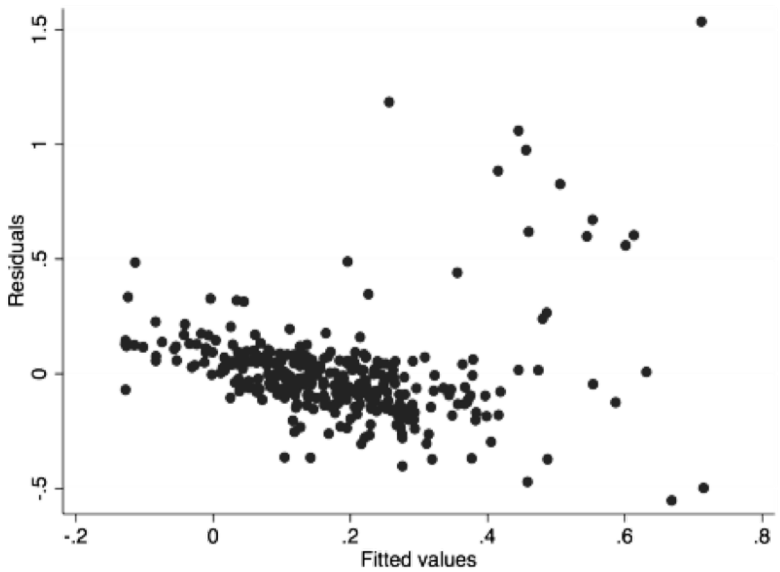

Figure 3. ***Heteroskedastic residuals in OLS regression.*** Residuals plotted against fitted values of an OLS regression on the monthly average of IPO initial returns from 1985 to 2020. In this regression, non-tech VC is used as an explanatory instead of the VC-variable. The figure illustrates the heteroscedastic errors and the necessity for performing weighted least-squares regressions.

The first thing to note in Table V is the superiority of the WLS (MLE) model over the OLS model at describing the data. For both columns without the bubbly dummy, column (aa) and (ba), their log-likelihoods increase by more than 100% from 38,712 to 86,490 and 37,529 to 86,490 respectively when going from OLS to WLS (MLE). The regressions that contain a bubble dummy variable also see an improvement in fit when modeled by WLS (MLE). However, their fits do not improve as dramatically as the specifications without the bubble dummy do, but instead go from a log-likelihood of 139,399 to 154,432 using the VC-variable in column (ab) and 139,180 to 156,245 using the non-tech VC-variable in column (bb). This is because the bubble period is now captured by an explanatory variable, rather than by the adjustments performed by the WLS in the columns without the bubble dummy.

Going from Average IR (initial returns) in the WLS (MLE) in the VC column (aa) to Average IR in the WLS (MLE) in the non-tech VC column (ba), one can observe how the Tech-variable increases in magnitude slightly, from 0,234 to 0,252. This is an anticipated result as Tech in column (ba) now captures the positive effect VC had in (aa), which is the only variable that has been removed. The coefficients are in line with our hypotheses for Tech and VC, although we expected the two to be more similar in magnitude. Despite being insignificant and not very strong, the coefficients of non-tech VC-variable is negative and in line with our hypothesis. This could indicate that the Tech-variable in fact assumes a more correct value in (ba). However, these findings also raise questions regarding the true coefficient of Tech, as it is not clear whether the strong positive initial returns amongst tech firms, many of which are venture capital-backed, are attributable to Tech or if they may in fact be attributable to VC.

On aggregate, the removal of the VC-variable and the addition of the non-tech VC variable does not affect the explanatory values of the models (going from columns (aa, ab) to columns (ba, bb)) as the log-likelihoods remain essentially identical for the corresponding columns. Since the non-tech VC-variable is weak and insignificant, as observed in Table IV, it can be concluded that the VC-variable never added explanatory value to begin with. Thus, the entire model should be modeled better without the variable. This was confirmed after performing an identical regression as in column (aa) but without the VC variable: although the log-likelihood dropped from 86,490 to 85,90, which is because we removed a variable, the AIC, which penalizes for using unnecessary variables and added complexity, improved from -148,98 to -151,583. Although a very minor improvement, the VC variable can be deemed unhelpful in explaining the behavior of IPO initial returns in this model. This additional regression is not displayed in Table V.

Table V

Regressions on the average initial return and variance of IPO initial returns with firm-specific characteristics as independent variables

Two types of cross-sectional regressions are performed for four different specifications. The columns labeled OLS display ordinary least-squares regressions on the monthly average IPO initial returns. The parentheses show the one-sided t-statistics for each coefficient. The columns labeled MLE show maximum likelihood-estimations on the coefficients calculated through weighted-least squares regressions on the monthly average initial returns, with the monthly standard deviation of residuals as weights. The log of the variance of the residuals is assumed to be linearly related to our firm-specific characteristics. The parentheses show the one-sided z-statistics for each coefficient.

	1985 - 2020											
ecification	(aa)			(ab)			(ba) Non-tech VC			(bb) Non-tech VC		
		MLE			MLE			MLE			MLE	
	OLS	Average IR	Log(var) of IR	**OLS**	Average IR	Log(var) of IR	**OLS**	Average IR	Log(var) of IR	**OLS**	Average IR	Log(var) of II
tercept	-0,102	0,035	-6,349	0,138	0,138	-6,128	-0,036	0,085	-5,395	0,150	0,155	-5,360
	-(0,87)	(0,52)	-(8,04)	(1,58)	(2,26)	-(7,54)	-(0,36)	(1,21)	-(6,25)	(1,98)	(2,48)	-(6,09)
og(shares)	0,002	0,002	0,075	-0,004	-0,003	0,058	0,000	0,000	0,034	-0,004	-0,003	0,041
	(0,37)	(0,49)	(1,50)	-(0,89)	-(0,81)	(1,16)	-(0,05)	-(0,06)	(0,68)	-(1,03)	-(1,00)	(0,82)
ech	0,395	0,234	3,606	0,156	0,161	1,041	0,508	0,252	3,815	0,191	0,147	0,790
	(4,10)	(3,28)	(4,28)	(2,39)	(2,85)	(1,36)	(4,96)	(4,44)	(8,19)	(3,12)	(3,29)	(1,45)
C	0,201	0,060	0,751	0,056	0,008	0,347						
	(1,57)	(0,79)	(0,81)	(0,60)	(0,12)	(0,44)						
on-tech VC							-0,166	-0,057	-1,779	-0,007	-0,022	-3,274
							-(1,19)	-(0,50)	-(1,11)	-(0,07)	-(0,28)	-(2,27)
YSE	-0,052	-0,096	-1,467	-0,143	-0,133	0,471	-0,088	-0,120	-1,847	-0,149	-0,138	0,040
	-(0,67)	-(2,24)	-(3,25)	-(2,38)	-(3,04)	(0,85)	-(1,29)	-(2,67)	-(3,63)	-(2,75)	-(3,12)	(0,07)
ice update	0,631	0,211	2,463	0,206	0,147	0,294	0,659	0,207	2,777	0,211	0,165	0,494
	(2,30)	(2,18)	(2,78)	(1,07)	(1,65)	(0,33)	(2,29)	(2,06)	(3,40)	(1,05)	(1,86)	(0,58)
ubble				0,648	0,658	2,648				0,651	0,673	2,666
				(6,99)	(11,75)	(11,41)				(6,94)	(12,12)	(11,96)
-squared	0,322			0,625			0,318			0,625		
og likelihood	38,712	86,490		139,399	154,432		37,529	86,490		139,180	156,245	
IC	-65,425	-148,98		-264,798	-280,865		-63,058	-148,98		-264,36	-284,489	
ample size		7312			7312			7312			7312	

The predicted values for the average and standard deviation of IPO initial returns, from the WLS (MLE) in column (ba), are depicted in Figure 4 and Figure 5 respectively. At a glance, the model seems to predict the standard deviation of initial returns (Figure 5) better than it predicts the average of initial returns (Figure 4). For our final section, section III, we hope to improve the prediction of the average initial returns by incorporating autoregressive components.

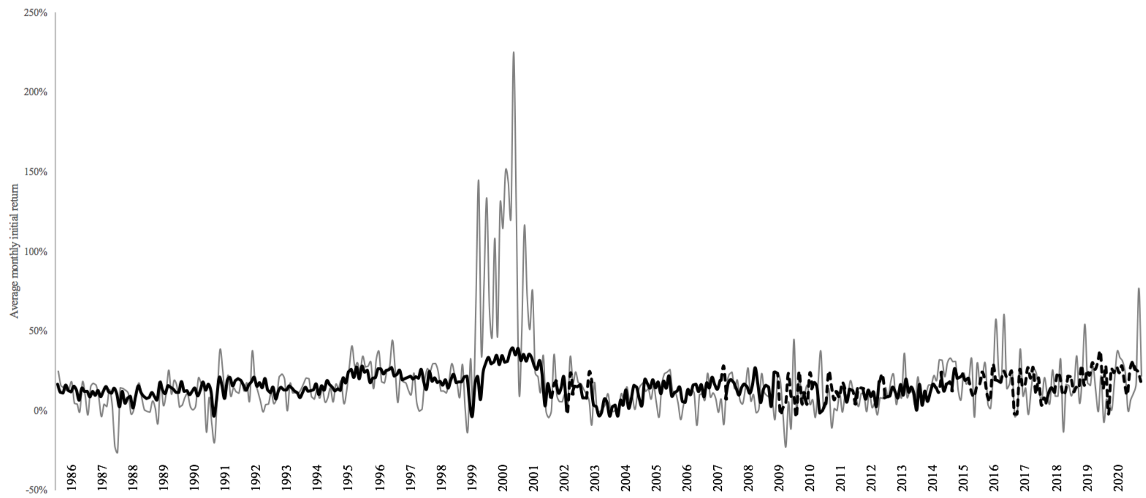

Figure 4: ***Observed versus predicted (by WLS (MLE)) monthly average IPO initial return.*** The gray line represents the observed monthly initial returns of IPOs, calculated as the average of that month's companies' percentage difference between their offer price and the stock price four weeks after taken public. The black line represents the monthly initial returns predicted by the MLE model in Table V column (ba). The dotted segments of the black line represent periods when there is insufficient data for at least one firm-specific characteristic, for which such values have been extrapolated by performing a weighted least-squares regression, using the variable in question as the dependent variable and using the remaining firm-specific characteristics with complete data as independent variables.

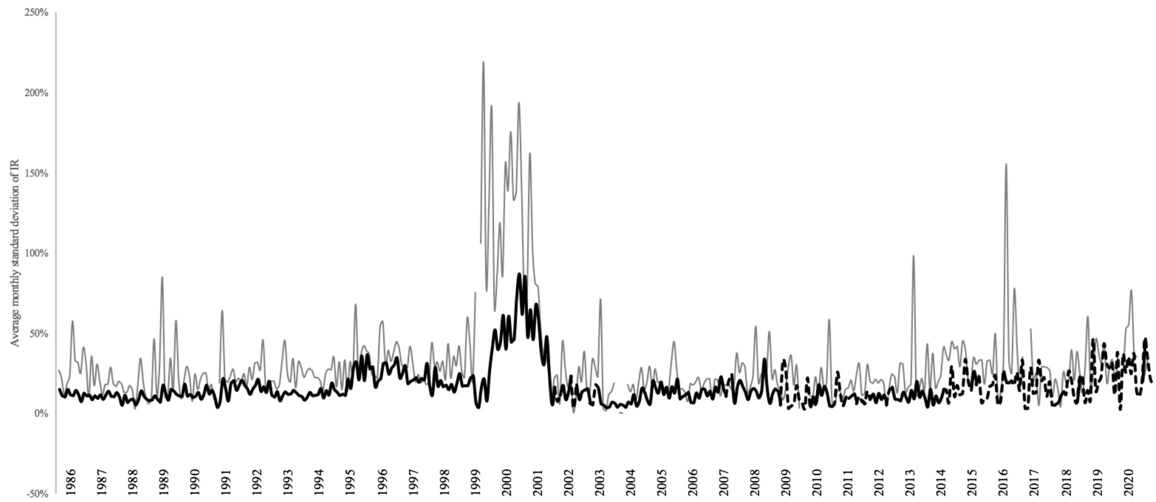

Figure 5: Observed versus predicted (by WLS(MLE)) monthly standard deviation of IPO initial returns. The gray line represents the observed monthly standard deviation of IPO initial returns, calculated as the standard deviation of that month's companies' initial returns. The black line represents the monthly standard deviation of initial returns predicted by the WLS (MLE) model in Table V column (ba). The dotted segments of the black line represent periods when there is insufficient data for at least one firm-specific characteristic, for which such values have been extrapolated by performing a weighted least-squares regression, using the variable in question as the dependent variable and using the remaining firm-specific characteristics with complete data as independent variables.

So far, the best model to describe the dataset was found to be column (bb) in Table V, with the non-tech VC variable and the dummy variable. However, to allow for the bubble period to be captured as a time-series phenomenon instead of as a variable, we will proceed with using WLS (MLE) in column (ba) as model for further analysis.

III. Time-Series Analysis

Clustering in firms with similar firm-specific characteristics, and the notion that market uncertainty lags in time and causes uncertainty on periods afterwards, leads one to expect time-series patterns in the average and volatility of IPO initial returns. In IPO pricing, a lack of recent and price-accurate listings should impair the underwriters' ability to price IPOs immediately afterwards too. This section serves to measure to what extent the average and volatility of IPO initial returns are dependent on their own past values.

To measure this, we have conducted two different time-series regressions; one ARMAX(1,1) process and one combined ARMAX(1,1) and ARCH(2) process. The prior will model the residuals as a function of past residuals, and the latter will model both the residuals and their variance as functions of past residuals. As we observed strong autocorrelation in our sample in Table II, which was not adequately modeled for in our WLS (MLE) or OLS regressions (e.g. see bubble period in figure 4 and 5), our hypothesis is that these operations should provide models that better fit the data. For comparison, the WLS models from columns (aa) and (ba) are included.

Table VI

Maximum likelihood-estimations on the mean and variance of IPO initial returns as time-series processes

The columns labeled MLE show maximum likelihood-estimations on the coefficients calculated through weighted-least squares regressions on the monthly average of IPO initial returns, with the monthly standard deviation of residuals as weights. The log of the variance of the residuals is assumed to be linearly related to the firm-specific characteristics (the coefficients for the firm-specific characteristics on variance are not displayed). The parentheses show the one-sided z-statistics for each coefficient. The Ljung-Box Q-statistics are based on the 20th lag of the autocorrelation function of the residuals, or the squared residuals, and has an asymptotic chi-2 distribution under the hypothesis of no autocorrelation. The p-values are displayed in parentheses below the Q-statistics. The columns labeled ARMAX(1,1) add an autoregressive term (AR) and a moving-average term (MA) to the error terms to correct for autocorrelation in the residuals. The columns labeled ARMAX(1,1) ARCH(2) add an autoregressive term to the variance of the residuals.

	1985 - 2020					
Specification	(a)			(b) Non-tech VC		
	MLE	**ARMAX(1,1)**	**ARMAX(1,1) ARCH(2)**	**MLE**	**ARMAX(1,1)**	**ARMAX(1,1) ARCH(2)**
	Average IR	Average IR	Average IR	Average IR	Average IR	Average IR
Intercept	0,035	0,132	0,310	0,085	0,152	0,304
	(0,52)	(0,95)	(7,04)	(1,21)	(1,04)	(6,73)
Log(shares)	0,002	0,002	-0,010	0,000	0,002	-0,010
	(0,49)	(0,31)	-(4,10)	-(0,06)	(0,26)	-(4,03)
Tech	0,234	0,039	0,217	0,252	0,078	0,203
	(3,28)	(0,34)	(3,81)	(4,44)	(0,75)	(4,13)
VC	0,060	0,082	-0,025			
	(0,79)	(0,64)	-(0,41)			
Non-tech VC				-0,057	0,006	-0,001
				-(0,50)	(0,03)	-(0,01)
NYSE	-0,096	-0,144	-0,292	-0,120	-0,157	-0,290
	-(2,24)	-(1,81)	-(8,89)	-(2,67)	-(1,73)	-(8,06)
Price update	0,211	0,194	0,113	0,207	0,202	0,112
	(2,18)	(1,93)	(1,60)	(2,06)	(2,14)	(1,56)
AR		0,917	0,894		0,918	0,893
		(57,29)	(31,33)		(59,85)	(30,57)
MA		-0,571	-0,558		-0,572	-0,557
		-(16,18)	-(13,26)		-(16,16)	-(13,05)
	Variance	Variance	Variance	Variance	Variance	Variance
ARCH intercept			0,009			0,009
			(7,69)			(7,74)
ARCH			0,8677			0,857
			(6,65)			(6,70)
Ljung-Box Q-statistic (20 lags)	617,05	49,64	42,60	632,16	49,57	42,65
(p-value)	(0,000)	(0,000)	(0,002)	(0,000)	(0,000)	(0,001)
Ljung-Box Q-statistic (20 lags, squared residuals)	398,67	277,44	261,33	395,37	273,86	263,49
(p-value)	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)
Log-likelihood	86,49	110,24	189,93	86,49	109,82	189,81
AIC	-148,98	-202,47	-359,86	-148,98	-201,64	-359,53
Sample size	7312	7311	7311	7312	7311	7311

In line with Lowry et al. (2010), our AR-terms are all close to one, indicating a positive lag effect on the average of IPO initial returns. For example, this means a positive residual of 1‰ at time t gives a positive residual of 0,9‰ at time t+1. In contrast to Lowry et al. (2010), whose model gave a similar MA-term of close to one, our MA-terms all took on negative values around -0,6. This means a positive error term of 1‰ in the residuals at time t gives a negative error term in the residuals of 0,6‰ at time t+1. In combination, the terms yield a positive but diminishing lag effect on the average of IPO initial returns. The coefficients are strongly significant and the positive autocorrelation in the residuals is captured effectively. This can be seen in the Ljung-Box Q-statistics, which drop dramatically for both columns (a) and (b) when going from MLE, to ARMAX(1,1), to ARMAX(1,1) ARCH(2). The log-likelihoods and AIC scores also improve. The prediction performed by the ARMAX(1,1) ARCH(2) model with the non-tech VC-variable is illustrated in Figure 6, which shows a remarkable improvement compared to the WLS (MLE) in Figure 4.

For the MLE model in column (b), only Tech, NYSE, and Price Update had coefficients that were significantly different from zero (|z-statistic|>1,96). Introducing the ARMAX(1,1) model reduced the strength of the coefficients further. However, when going to the ARMAX(1,1) ARCH(2) model, all coefficients improved and only non-tech VC and Price Update remained insignificant. Although the non-tech VC-variable consistently has been insignificant in previous sections, the Price Update variable seems to have mistakenly captured time-dependency in the variance of the residuals in previous regressions.

Another intriguing observation in column (a) is that the VC-variable shifts sign and remains insignificant when moving from MLE and ARMAX(1,1) to ARMAX(1,1) ARCH(2). Although never significantly so, the coefficients of VC have consistently been positive previously, and this finding could give an indication that the true coefficient of VC is not positive, as priorly thought. The non-tech VC-variable has been negative throughout both section II and III, and the convergence of both variables in the ARMAX(1,1) ARCH(2) model suggests that the VC variable may not only have captured the effect of risky industries, but also the effect of conditional variance in the residuals.

Finally, it is worth remarking that the models using the VC-variable versus the models that use the non-tech VC-variable are nearly identical in other coefficients, log-likelihoods, AICs, and Ljung-Box scores. In Lowry et al. (2010), the Tech-coefficients remain roughly constant throughout their corresponding three models, around 0,075. Our tech in both ARMAX(1,1) regressions are consistent with this, but in our MLE and ARMAX(1,1) ARCH(2) models with the VC-variable, our Tech increases to 0,234 and 0,217 respectively. A similar pattern is observed in column (b) with the non-tech VC variable. The reason for this result is unclear.

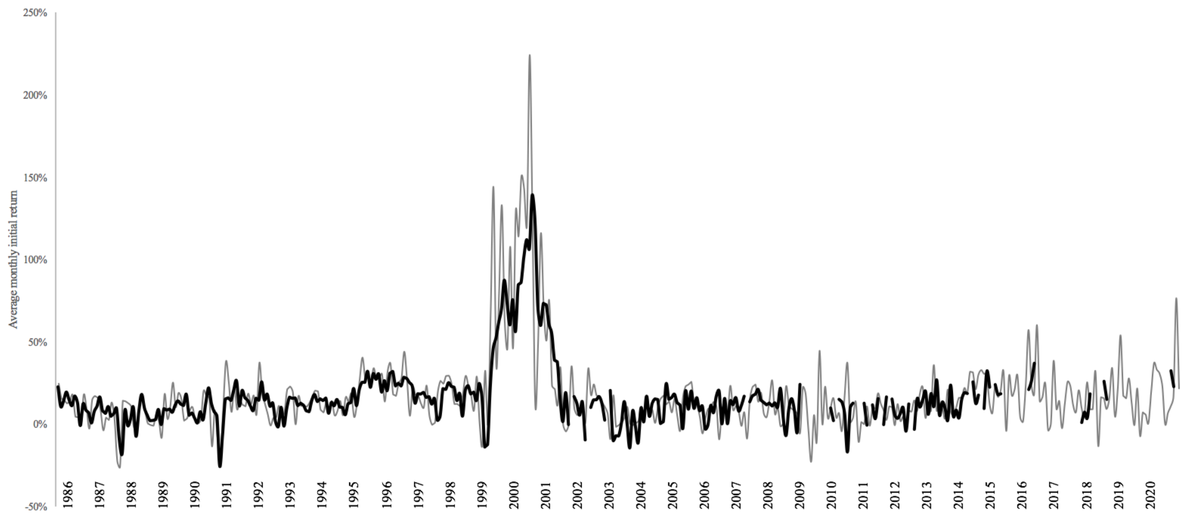

Figure 6: ***Observed versus predicted (with ARMAX(1,1) ARCH(2)-process) monthly average IPO initial return.*** The gray line represents the observed monthly initial returns of IPOs, calculated as the average of that month's companies' percentage difference between their offer price and the stock price four weeks after taken public. The black line represents the monthly initial returns predicted by the ARMAX(1,1) ARCH(2) model in Table VI, column (b). Compared with Figure 4, which displays predicted values generated by a WLS (MLE), the new prediction shows an illustrative improvement in fit, which is confirmed by the log-likelihoods and AICs (see Table VI column (b), MLE versus ARMAX(1,1) ARCH(2)).

Conclusions

This study serves as a complementary study to the premier model on IPO underpricing to date, developed by Lowry et al. (2010), but with a particular focus on venture capital-backing. It has been done on both previously studied data from the period 1985 to 2005 and on hitherto unexplored data from 2006 to 2020. Following the methodology of Lowry et al. (2010), we have explored firm-specific factors and time-dependent components as underlying factors driving IPO underpricing and initial return volatility. The effect of venture capital-funding on IPO underpricing is of particular importance since it is one of few factors entrepreneurs and investors can act upon (choose to accept VC-funding or not) and thus prevent their firm from going public at below market value. Additionally, VC-funding has increased on average in recent years (see appendix A) and it is therefore increasingly important to provide conclusive evidence of its true effect, which is not agreed upon in literature. It is most suitable to perform this analysis on the model of Lowry et al. (2010) as it is the most sophisticated model yet; particularly as they suspected that their VC variable picked up a "risky industry effect".

We show that the VC-variable lacks explanatory power. To some extent, this was demonstrated by the increased explanatory value (decrease in AIC score) that was achieved when estimating a weighted least-squares regression without the variable in question as this means the VC-variable solely added complexity to the model. However, the true effect of venture capital-backing is still uncertain. The coefficient of our non-tech VC-variable is close to zero and negative throughout all sections, which the VC-coefficient eventually became as well in the ARMAX(1,1) ARCH(2) regression. However, these coefficients were insignificant we are only given an indication of the effect of venture capital-backing, which is that it likely contributes less to underpricing than previously thought. Finally, by replacing VC with non-tech VC, we effectively eliminated the multicollinearity between Tech and VC, thus improving the specification and usability of the model for when more data is available.

Compared to previously explored time periods (except when including the bubble period), the period from 2006 to 2020 exhibited higher IPO underpricing, higher volatility of initial returns, and a higher correlation between the two. Despite this, almost all correlations between firm-specific characteristics and the average and standard deviation of IPO initial returns decreased (see Table IV). In combination with the period's lower autocorrelations, as seen in Table II, this suggests there may be other variables beside firm-specific characteristics and time-contingency that better explain IPO underpricing in recent years.

In conclusion, our findings suggest that entrepreneurs and investors should not necessarily see venture capital-funding as disadvantageous for the success of a future IPO, as suggested the results of Lowry et al. (2010). On the contrary, it appears as if the involvement of a venture capital-partner indeed does mitigate underpricing, although to what extent this is the case remains unknown.

Directions for further research

The focus of this study has been to examine the monthly level and dispersion of IPO initial returns from 1985 to 2020 and delving deeper into how different firm-specific characteristics have an impact on the difficulty of accurately pricing an IPO. We find that the venture capital variable in previous findings does not yield a meaningful interpretation, either by itself or in the model as a whole.

For our dataset, a clear majority of firms that receive support from venture capital firms operate in high-tech industries: a variable that in itself contributes to uncertainty in the valuation process. As a consequence, a lot of observations are lost when constructing a non-tech VC-variable. The sole strong conclusion we can draw is that it is unfavorable to model the VC-variable and the Tech-variable jointly, as one only adds complexity in the presence of the other. However, we believe our findings validate further research on this subject. With more observations on venture capital-backed firms outside high-tech industries, one could perhaps isolate the true effect of venture capital funding on IPO underpricing.

Moreover, as seen in the values in Table IV, all coefficients of firm-specific characteristics take on lower values in our extended time set and no variable has increased or decreased in relative magnitude or importance. At the same time, the mean and cross-sectional standard deviation of initial returns have increased for our time extension. This is not explained by increased time-dependency in underpricing either, as can be seen in Table II. Altogether, this creates the impression that other variables may be of larger importance today compared to before 2006. A topic for further research would therefore be to concretize and assess factors beyond firm-specific characteristics and time-dependence as explanatory variables for IPO underpricing. Market-wide volatility in equity prices, as examined by Lowry et al. (2010), should also be considered as one such factor.

If one were to deviate from the framework developed by Lowry et al. (2010), another approach would be to ungroup the data points and revert them back into individual companies (at the sacrifice of volatility measurements). This way, one would possibly see the true effect of VC as the number of observations increase.

References

Beatty, R. P., & Ritter, J. R. (January-February 1986). Investment banking, reputation, and the underpricing of initial public offerings. *Journal of Financial Economics, 15*(1-2), pp. 212-232.

Box, G. E., & Jenkins, G. M. (1976). *Time series analysis: forecasting and control, rev. ed.*

Dong, M., Michel, J.-S., & Pandes, A. J. (2011). Underwriter Quality and Long-Run IPO Performance. *Financial Management, 40*(1), pp. 219-251.

Engle, R. F. (July 1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica, 50*(4), pp. 987-1007.

Francis, B. B., & Hasan, I. (April 2001). The Underpricing of Venture and Nonventure Capital IPO:s An Empirical Investigation. *Journal of Financial Services Research, 19*, pp. 99-113.

Guiso, L. (March 1998). High-tech firms and credit rationing. *Journal of Economic Behaviour & Organization, 35*(1), pp. 39-59.

Hanley, K. W., Kumar, A. A., & Seguin, P. J. (October 1993). Price stabilization in the market for new issues. *Journal of Financial Economics, 34*(2), pp. 177-197.

Kohers, N., & Kohers, T. (April 2004). Information sensitivity of high-tech industries: evidence from merger announcements. *Applied Financial Economics, 14*(7), pp. 525-536.

Lee, P. M., & Wahal, S. (August 2004). Grandstanding, certification and the underpricing of venture capital backed IPOs. *Journal of Financial Economics, 73*(2), pp. 375-407.

Ljung, G. M., & Box, G. E. (August 1978). On a measure of lack of fit in time series models. Biometrika, 65(2), pp. 297-303.

Ljungqvist, A. (2007). Chapter 7 - IPO Underpricing. i *Handbook of Empirical Corporate Finance SET.* Elsevier B.V.

Loughran, T., & Ritter, J. (2004). Why Has IPO Underpricing Changed Over Time. *Financial Management, 33*(3), pp. 5-37.

Lowry, M., Officer, M. S., & Schwert, W. G. (April 2010). The Variability of IPO Initial Returns. *The Journal of Finance, 46*(2), pp. 425-465.

Megginson, W. L., & Weiss, K. A. (July 1991). Venture Capitalist Certification in Initial Public Offerings. *The Journal of Finance, 46*(3), pp. 879-903.

PwC; CB Insights. (January 2020). *Statista.* Taken from https://www-statista-com.ez.hhs.se/statistics/277501/venture-capital-amount-invested-in-the-united-states-since-1995/

Rock, K. (January-February 1986). Why new issues are underpriced. *Journal of Financial Economics, 15*(1-2), pp. 187-212.

Ruud, J. S. (October 1993). Underwriter price support and the IPO underpricing puzzle. *Journal of Financial Economics, 34*(2), pp. 135-151.

Appendix

A. Value of venture capital investment in the United States from 1995 to 2019