Charles University in Prague
Faculty of Social Sciences
Institute of Economic Studies

BACHELOR THESIS

The effect of introduction of Cloud Computing: The case of Venture Capital

Author: **Jan Šomvársky**

Supervisor: **prof. Ing. Michal Mejstřík CSc.**

Academic Year: **2015/2016**

Declaration of Authorship

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain a different or the same degree.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, 12th May 2016

Signature

Acknowledgments

I would like to express my sincere gratitude to prof. Ing. Michal Mejstřík CSc. for supervising my thesis and his support.

I am also grateful to PhDr. Jaromír Baxa Ph.D. and Bc. Lenka Röhryová for their invaluable comments.

Abstract

The thesis aims to examine the impact of introduction of Cloud Computing on Venture Capital (VC) financing in the United States. In the first part we review features of Cloud Computing and their impact on startup costs in context of VC. In this thesis we consider Amazon Web Services (AWS), introduced in 2006, a pioneer of widely accessible Cloud Computing. In the second part we quantify the cost reduction associated with utilization of AWS against owning IT infrastructure. Results show 529 fold decrease in startup costs in 3-month time frame. In the third part we analyze the impact of introduction of AWS on seed and later-stage investments in context of selected macroeconomic and technological factors. We perform analysis on a comperhensive dataset from National Venture Capital Association using Autoregressive Distributed Lag (ARDL) model to account for a change in lagged values of dependent and independent variables. Main finding of our analysis suggests that seed investments are significantly influenced by the introduction of AWS and subsequent drop in startup costs. Specifically, the decline in cost of startup induced 29.67% increase in seed investments. Further findings indicate insignificant relationship between seed investments and macroeconomic factors. Moreover, according to our results, later-stage investments show no significant relationship with introduction of AWS and associated cost reduction. The outcome of the thesis is that introduction of AWS and subsequent startup cost reduction in 2006 significantly increased seed investments but did not influence later-stage investments. Moreover, seed investments are not significantly affected by macroeconomic factors.

JEL Classification	G24, M13, M15, O31,
Keywords	cloud computing, amazon web services, elastic compute cloud, venture capital, seed investments

Author's e-mail	`somvarsky@gmail.com`
Supervisor's e-mail	`mejstrik@fsv.cuni.cz`

Abstrakt

V této bakalářské práci zkoumáme vliv Cloud Computingu na rizikový kapitál v USA. V první části práce shrnujeme hlavní přednosti Cloud Computingu a odhadujeme velikost nákladů, které tyto společnosti ušetří oproti společnostem, které vlastní svojí IT infrastrukturu. Za průkopníka tohoto směru považujeme Amazon Web Services (AWS), který zpřístupnil Cloud široké veřejnosti. Ve druhé části odhadujeme snížení nákladů spojených s využitím AWS oproti vlastnictví IT infrastruktury. Výsledky ukazují 529 násobné snížení nákladů v prvních 3 měsících činnosti společnosti. Tuto změnu v nákladech potřebných na start společnosti zkoumáme v kontextu rizikového kapitálu, makroekonomických proměnných, a vybraných technologických faktorů. Pro ekonometrickou analýzu volíme data z National Venture Capital Association a Autoregressive Distributed Lag (ARDL) model pro jeho schopnost zachytit efekt zpoždění ve vysvětlující i vysvětlované proměnné. Hlavní výsledek této práce je, že spuštění AWS a snížení nákladů na spuštění společnosti nepřímo úměrně ovlivňuje celkový objem seed investic. Konkrétně propad v nákladech potřebných na spuštění společnosti způsobený osvojením AWS zapříčinil 29,67% nárůst celkového objemu seed investic. Další zkoumání naznačuje, že seed investice nejsou signifikantně závislé na makroekonomických faktorech. V případě investic v pozdějších fázích společnosti jsme nenašli žádnou závislost na změně v nákladech na spuštění. Výsledek této práce naznačuje, že AWS v roce 2006 způsobil významný nárůst v celkovém objemu seed investic a investice v pozdějších fázích společnosti naopak neovlivnil. Navíc seed investice nejsou signifikantně ovlivněny makroekonomickými faktory.

Klasifikace JEL	G24, M13, M15, O31,
Klíčová slova	cloud computing, amazon web services, elastic compute cloud, rizikový kapitál, seed investice

E-mail autora	`somvarsky@gmail.com`
E-mail vedoucího práce	`mejstrik@fsv.cuni.cz`

Contents

List of Tables

List of Figures

Acronyms

VC Venture Capital

S3 Simple Storage Service

EC2 Elastic Compute Cloud

AWS Amazon Web Services

NVCA National Venture Capital Association

IPO Initial Public Offering

NIST National Institute of Standards and Technology

OPEX Operational Expenses

CAPEX Capital Expenses

ARDL Autoregressive Distributed Lag

ADF Augumented Dickey-Fuller

AIC Akaike Information Criterion

BIC Bayiesian Information Criterion

RD Research&Development

GDP Gross Domestic Product

ACF Autocorrelation Function

IT Information Technology

OLS Ordinary Least Squares

Bachelor's Thesis Proposal

Author	Jan Šomvársky
Supervisor	prof. Ing. Michal Mejstřík CSc.
Proposed topic	The effect of introduction of Cloud Computing: The case of Venture Capital

Topic characteristics

The aim of this thesis is to examine the impact of introduction of Cloud Computing on Venture Capital (VC) financing in the United States. During year 2006, Amazon introduced Simple Storage Service (S3) and Elastic Compute Cloud (EC2) services which allow companies to utilize Amazon's computing capabilities on a pay-as-you-go basis. This innovative business model enabled entrepreneurs to start companies without the requirement of building own IT infrastructure. This dramatically reduced upfront costs needed to setup an Internet-related company. Consequently, barriers to entry to Internet-related industries fell significantly, which opened new opportunities for entrepreneurs and VC investors. We presume that it led to increase in total seed financing. We also presume that introduction of S3 and EC2 did not influence later-stage investments as developed companies already have their computing capabilities.

We examine dataset published by NVCA in cooperation with Pricewaterhouse Coopers based on data from Thomson Reuters VentureXpert. It contains US VC investments from 1995 till 2015 and we analyze it using Autoregressive Distributed Lag (ARDL) model. We complement this dataset by development of cost of IT capabilities, macroeconomic factors, selected technological variables that are potent to influence seed investments, and government subsidies to small and medium entreprises.

The outcome of this thesis determines statistical and economical significance of introduction of Amazon's Cloud Computing services on VC investment activity in USA.

Hypotheses

1. Negative shock in cost of starting a company positively influences seed investments.
2. Macroeconomic factors do not influence seed investments.
3. Shock in cost of starting a company does not influence later-stage investments.

Outline

1. Introduction
2. Literature review
3. Estimation of IT cost reduction
4. Methodology and Data
5. Results
6. Conclusion

Core bibliography

1. Ewens, M., R. Nanda, & M.Rhodes-Kropf (2015): "Cost of experimentation and the evolution of venture capital." *Harvard Business School Entrpreneurial Management Working Paper* **(15-070)**.
2. Greene, W. H. (2003): "Econometric analysis, Volume 5." *Prentice hall.*
3. Koomey, J. G., C. Belady, M. Patterson, A. Santos, & K.-D. Lange (2009): "Assessing trends over time in performance, costs, and energy use for servers." *Lawrence Berkeley National Laboratory, Stanford University, Microsoft Corporation, and Intel Corporation, Tech. Rep.*
4. Gompers, P. & J.Lerner (2001): "The venture capital revolution." *The Journal of Economic Perspectives* **15(2)**: pp. 145-168.
5. Zhang, Q., L. Cheng, & R.Boutaba (2010): "Cloud computing: state-of-the-art and research challenges." *Journal of internet services and applications* **1(1)**: pp. 7-18.

Author Supervisor

Chapter 1

Introduction

In this thesis, we examine the impact of change in costs of starting a company on Venture Capital (VC) investment activity in the United States. Samuel Kortum (2000) claims that VC has strong positive effect on innovation and is an important driver of economic productivity. The United States are the front-runner in this field with long history and the biggest market share in global. US VC activity accounted for $52.1bn in 2014 compared to China, the second biggest market, with only $15.5bn. (EY 2015) The importance of venture capital to economic prosperity can be illustrated by Gornall & Strebulaev (2015) who attempted to research the economic impact of VC-backed public companies in the US. According to this paper, 17% of all public companies are VC-backed, however, those VC-backed companies are responsible for 44% in spending on R&D. Furthermore, three[1] of top-five companies by market capitalization were developed by raising capital in form of VC.

In the past two decades, Internet heavily influenced most of the industries and caused lots of turbulences in VC activity. The most striking event occured around year 2000. In this year, total invested amount topped $104 998M which is 5-fold increase from year 1998. The originator of this huge leap are investments in Internet-related companies. Over the two following years total investments plummeted back to the level of 1998. This period is called the Internet bubble. Since the crash, VC industry has recovered and was rising constantly with relatively minor decrease in 2009 caused by global financial crisis. From 2013, Internet-related investments have been flourishing again ac-

[1] Apple, Google, and Microsoft

counting for about 68% of total investments in 2015. This suggests that the Internet became crucial part of entrepreneurship and VC.

Both entrepreneurial and VC environment adapted in a fundamental way as the Interned is reshaping the world. Over the course of past two decades, cost of starting an Internet venture decreased substantially mainly due to falling cost of computing power. One of the most significant events which affected cost of IT infrastructure was introduction of Amazon's cloud computing branch called Amazon Web Services (AWS). This solution enabled its customers to rent computing capabilities without the requirement of acquiring costly hardware. This dramatically decreased cost of startup and encouraged grasp of new entrepreneurial opportunities. As VC seed financing is primarly concerned with enhancing development of new entreprises and bringing new products to the market, AWS had potential to fundamentally affect this industry. In this paper, we examine relationship of VC investments and cost of starting a company influenced by introduction of AWS.

In Chapter 2, we review past research related to VC and cost of starting a company along with overview of Cloud Computing. In the next chapter, we present our estimates of development of costs associated with purchase of own IT infrastructure and compare it to the cost of AWS. In Chapter 4, we describe the dataset and our model and examine three hypotheses. Our key hypothesis is concerned with relationship of decreased cost of startup and influx of VC seed investments. In the second hypothesis, we test whether other macroeconomic and technological factors affect seed investments. Further, we test our last hypothesis which states that later-stage investments are not correlated with the change in cost. In the Chapter 5, we present our results and discussion.

Chapter 2

Literature review

2.1 Theoretical background

In order to fully understand how VC activity can be influenced by change in startup costs, we have to understand the VC cycle. In this thesis we will use US National Venture Capital Association's (NVCA) definition: "Venture capital firms are professional, institutional managers of risk capital that enable and support the most innovative and promising companies". According to Paul Gompers (2001), VC firms have developed as financial intermediaries providing capital to companies that would have difficulties in raising capital by traditional methods (e.g. bank loans). Such companies are typically in their earliest stages of development and are operating under conditions of extreme uncertainty in a rapidly changing environment. Uncertainty is further deepened by company's possession of only few tangible assets and high levels of information asymmetry between investee and investor.

Typically, during the VC cycle, investor and investee go through several stages of cooperation. Classification of VC financing stages is rather vague and overlaps frequently. In this paper, we use definition from NVCA (2016) which is in line with classification of our dataset. In the initial stage, a startup receives a seed or angel investment. In this period company is typically in existance for less than 1.5 years and is in phase of product development. In seed stage, investors usually aim to deliver product to general public. It is the riskiest stage as investors have limited validation of company's business model and team's

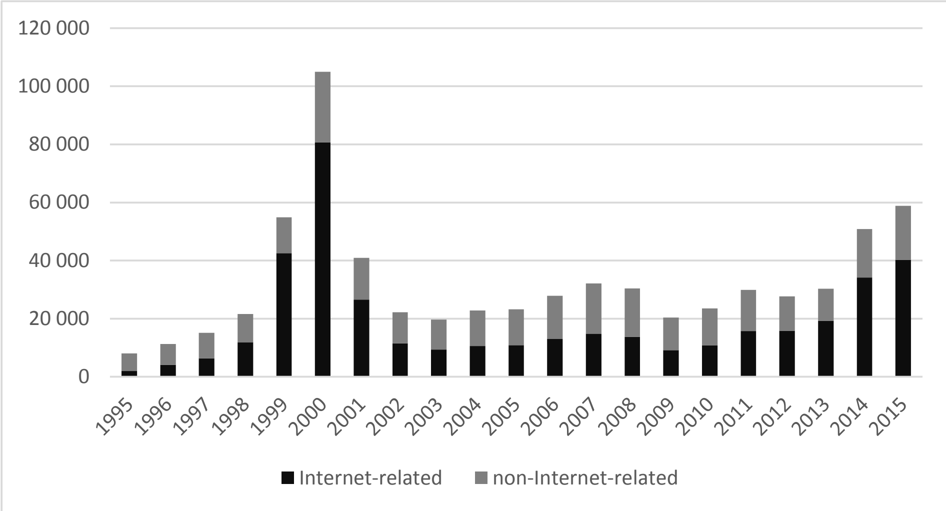

Figure 2.1: Total VC investments (in $ Millions)

executive abilities. It is also crucial period for the startup as the viability of business shows up. In this stage, investor is screening potential deals and filtering viable investment opportunities. According to prominent angel investor Ari Korhonen, the most significant determinants of whether to invest in an early-stage company are the management team, the idea, size of potential market, scalability and competitors' barriers to entry. (Korhonen 2016) Technically, a VC investment is exchanged for an equity which is illiquid until investee's maturity. Investors not only provide capital to investees but also participate on strategic planning and advisory by taking a board seat. Active engagement in a company's operations is crucial to company's success and occurs most intensively in early stages. (Bernstein *et al.* 2015) It is common that the investor takes part in everyday decisions during the first year which, naturally, limits the number of companies that the investor can fund.

In the following stage, called early stage, a company receives Series A or B funding and already has a product which not neccessarily generates revenue. This happens oftentimes when company is less than 3 years in operation. In expansion stage, company is generating revenue and is growing rapidly, however, it might not generate profit. The investment helps company to scale its business and become profitable. Later-stage investments are aimed at companies which have their product widely accessible and show profit. Such investments help companies further expand their business.

At investment's maturity, divestment is most frequently performed by initial public offering (IPO) or acquisition by another subject. This happens on average after 5 to 8 years from company's startup. (NVCA 2016)

Recently, the Internet hype gave rise to alternative methods of private-venture financing which largely contribute to democratization of investing in startups. The concept, which received the most attention in VC industry, is Equity Crowdfunding. This new approach is becoming an important intermediary in raising funds by startups. The novelty of this alternative approach brought up many regulatory issues. Government started to address those in 2013 when Equity Crowdfunding became legal to accredited investors in the US under JOBS act[1]. Status quo remained until October 2015 when Equity Crowdfunding was officially permitted for general public[2]. This was an important moment in VC industry as it enabled masses to take part in the Internet rush. Equity Crowdfunding investment volume reached $1.2bn in 2015 in the US and is expected to double in 2016[3].

2.2 Cost of starting a company

2.2.1 Era prior to modern Cloud Computing

Nowadays, we are experiencing true democratization of entrepreneurship. The Internet allows entrepreneurs to start a new venture without the requirement of owning almost any physical assets. In 1990', start-up costs in Internet-related industries were significantly higher due to the requirement of building own IT infrastructure. To start an Internet company, entrepreneurs had to buy servers, acquire costly software licenses and hire IT support staff. This involved very high upfront costs and substantial operational costs. It was accompanied by other serious drawbacks such as distraction from company's primary business. Cost of IT equipment fell significantly with the emergence of open-source software, which was pioneered in 1980' by Free Software Movement. Since then, entrepreneurs no longer had to buy expensive software bundled with hardware and could take advantage of community-developed solutions like Linux. Till 2006, IT expenditures and barriers to entry were still relatively

[1]https://www.sec.gov/spotlight/jobs-act.shtml
[2]https://www.sec.gov/news/pressrelease/2015-249.html
[3]http://crowdexpert.com/crowdfunding-industry-statistics

prohibitive. Companies had to guess expected bandwidth and adjust their infrastructure accordingly. It was vastly inefficient as the infrastructure had to be designed to withstand peak loads. Thus it was underutilized most of the time. Also, such kind of solution was rigid and took relatively long time to adjust for change in bandwidth. The above mentioned drawbacks disappeared with the introduction of Cloud Computing.

2.2.2 Cloud Computing and reduction of startup costs

Cloud Computing or On-demand Computing has recently changed the way how companies approach delivering services over the Internet. According to Zhang *et al.* (2010), Cloud Computing is a combination of a set of existing technologies that enables running businesses in a different way and it is meant to meet technological and economical requirements of today's demand for IT. The true innovation in On-demand Computing is not technology itself but the business model. According to National Institute of Standards and Technology (NIST) definition, "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction...". NIST further characterizes Cloud by 5 features:

On-demand self-service. User can exploit computing power such as server time and data storage according to his demand without requiring interaction with any human operating cloud solution.

Broad network access. Computing capabilities are delivered over the network and can be accessed by various thin or thick client platforms such as PCs and mobile phones.

Resource pooling. The cloud provider pools computing resources, such as computing power and data storage, to serve multiple users by dynamically assigning and reassigning virtual and physical resources in order to meet users' demand.

Rapid elasticity. Users are flexibly provisioned with computing capabilities to upscale or downscale rapidly according to their demand and without restrictive limitations.

Measured service. Cloud provider automatically measures and optimizes use of resources in order to maintain transparency and efficiency of the service.

Furthermore, cloud systems are broken down into 3 service models.

Software as a Service (SaaS). The user is provided with applications, which are running on a cloud infrastructure and are accessible through thin client platform or a program interface.

Platform as a Service (PaaS). The capability provided to the user is to deploy user-created applications created using tools supported by the provider. User has control over his applications and is able to configure them.

Infrastructure as a Service (IaaS). User is provided with fundamental computing resources to deploy and run any application created by the user. User does not manage cloud infrastructure but is able to control his applications and used resources.

With such solution, entrepreneurs can start an Internet venture with nearly zero upfront IT costs and optimal IT OPEX as they pay only for used bandwidth. If we consider that expenses on computing capabilities present considerable part of total costs required to start an Internet-related company, it suggests that barriers to entry fell significantly. The pioneer of modern Cloud Computing is Amazon with its Simple Storage Service (S3) and more importantly Elastic Compute Cloud (EC2), both introduced in 2006 under Amazon Web Services (AWS) branch. As names suggest, S3 is primarly concerned with data storage while EC2 aims to deliver computing power. Amazon aimed to utilize its excess computing capacity and provide data storage and computing capabilities at significantly reduced cost. Amazon's excess capacity was soon exhausted and AWS had to grew into an IT infrastructure provider with whopping capacity.

Nowadays, Cloud Computing is still in its infancy but it has already proven its tremendous impact on a number of related and nonrelated technologies. The most significant reduction in costs is seen in adoption by early-stage businesses where it can drastically reduce upfront expenditures and streamline whole IT process. The effect of Cloud Computing on established companies is ambiguous as it might be less costly to build own IT infrastructure in certain cases. As Williams (2012) states, "The reduction in overall CAPEX and OPEX associated with cloud computing diminishes the barriers to entry for new firms, enabling startups to have equal footing with established players in terms of computing power." Democratization of computing power enables rise in innovation and R&D. Williams (2012) for example emphasizes the ability of cloud computing to streamline development of new medicines. This can be illustrated by an example stated by Jason Stowe, CEO of Cycle Computing cloud company "For example, 30 000-core cluster for top-five pharma would have cost $5 to $10 million and about six months to build. With Cycle's cloud offering, the project took eight hours to implement, at a cost of about $10 000." [4] As recognized venture capitalist and serial entrepreneur Marc Andeerssen claimed "...they (today's startups) go on Amazon Web Services and they pay by the drink and they're paying somewhere between 100x and 1000x cheaper per unit - per unit of compute, per unit of storage, per unit of networking, per unit of software.". Amazon has changed Internet entrepreneurship in a fundamental way from owning IT infrastructure towards subscription model with maximal flexibility.

As Michael Grant, CEO of Cloudscaling argued in retrospective: "Amazon Web Services has been the biggest boon to venture capital-backed companies in recent years. It has meant you can now fund 10 companies for the price of one, and you are seeing new applications being developed that would have been difficult to build cost-effectively in the early days" [5]

To our knowledge, only Ewens *et al.* (2015) attempted to research the impact of technological shocks to financing of private-held firms. Their paper examined the impact of introduction of AWS on VC investment strategy. The methodology used in this paper involved difference-in-difference analysis of groups of companies which received funding in periods 2002-2006 and 2006-2010. The main finding shows shift towards "spray and pray" approach - providing less capital and governance to a higher number of startups which are likely to reveal their potential quickly and cheaply. It decreases cost of experimentation and alters VC investment strategy from investing to complex technologies towards investing in ventures which reveal their prospects cheaply and quickly.

[4] http://www.forbes.com/sites/joemckendrick/2011/11/01/cloud-computing-is-fuel-for-the-next-entrepreneurial-boom/#5acd80dd7070

[5] http://www.ft.com/cms/s/0/fc871bca-58e1-11e1-b9c6-00144feabdc0.html#axzz43YMeI4IN

Chapter 3

Estimation of IT cost reduction

In this chapter we will quantify change in IT expenditures with the introduction of AWS. Firstly, we will estimate cost development of entry-level IT infrastructure before year 2006 by assessing total cost of ownership for particular server cluster. Than, we will compare it with corresponding server configuration on AWS and construct time-series.

3.1 Total cost of ownership

We examined report TechWise (2004) which studies and compares three server clusters available on the market in 2004. TechWise developed methodical approach called Reliability-Adjusted Total Cost of Ownership, which takes into account not only hardware acquisition costs but also managerial and downtime costs over three-year period. TechWise defines downtime as "the number of hours per year, if any, when cluster's primary application(s) were not able for end-users to access." This can be caused for example by software viruses or hardware failure. The study was conducted by interviewing 94 IT professionals in US firms and comparing three entry-level and mid-range server clusters. The interview was focused to gain data on costs associated with installation, management and maintenance along with data on number of downtime hours and associated costs. Obtained dataset was complemented by actual pricing from IDEAS International and Reliability-Adjusted TCO was calculated.

Results revealed that four main cost drivers are (1) server acquisition and service contract, (2) cluster installation and configuration, (3) cluster management and maintenance, (4) costs associated with downtime. For entry-level clusters, the results showed that the highest costs over the three-year period are associated with management and downtime representing 47% and 44% respectively. Hardware acquisition and installation & training costs amount to 7% and 2% respectively. Conservative average downtime cost per hour for entry-level solutions was estimated by TechWise at $10 000. The outcome of the paper is that the most economical solution amongst 2-way clusters (basic cluster)[1] is HP AlphaServer DS 25 running OpenVMS. This solution costs $913 000 compared to Sun Solrais ($1 456 000) and IBM AIX ($1 348 000).

In order to produce more comparable results, we estimate simple Total Cost of Ownership (TCO) for the most economical solution based only on Acquisition, Installation & Training, and Management costs. If we break down the total costs associated with purchase and operation of HP AlphaServer cluster, the upfront expenditure represents the sum of Acquisition and Installation & Training which totals $82 170. Operational costs are represented by Management which totals $429 110 over the three-year period resulting in overall TCO of $511 280. In order to be able to compare costs in different phases of company's development, we calculate cumulative expenditures associated with building own IT infrastructure at startup. Those include only upfront costs, and costs 1, 3, and 12 months after company's launch. (See Figure 3.1) To make a comparison with AWS, we need to define the time frame. In time zero, reduction in expenses on AWS would go to infinity as using AWS involves negligible upfront costs. We use 3-month period which corresponds to frequency of our data and accurately represents costs in the earliest stage. Therefore, we assume startup costs in 2004 equal to sum of Acquisition, Installation & Training, and Management costs for 3 months of operation. The result for the above setting amounts to $117 929.

In order to determine development of IT infrastructure costs, we researched Koomey *et al.* (2009) who examines trends in development of cost drivers associated with purchase and operation of IT equipment. This study takes into account energy consumption and purchase costs. Expenditures are then adjusted for inflation and reported as real cost of server for years 2001, 2004 and

[1]equivalent configuration to 2 nodes with 2 processors per node, 2 GB of memory per node and 438 GB of storage array

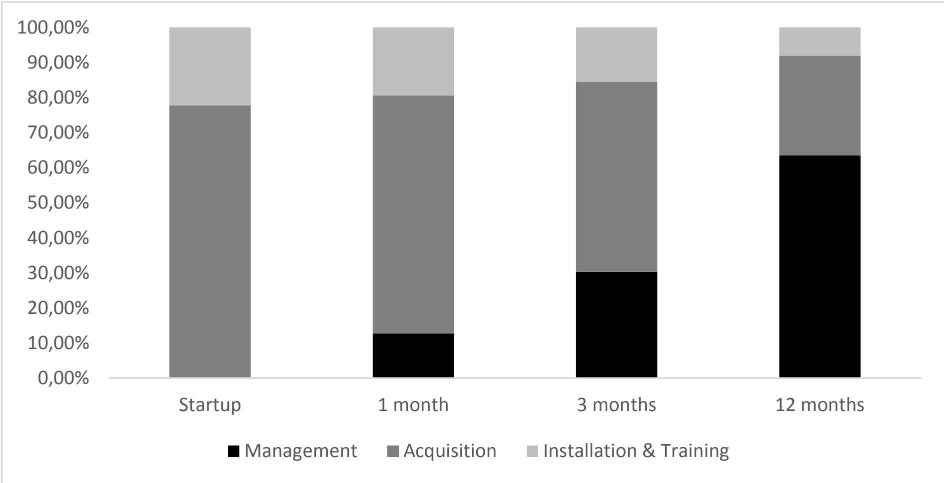

Figure 3.1: Breakdown of cumulative expenditures on HP AlphaServer cluster

2008. The results show that cost of 1 unit of generic server was 1 584, 1 482 and 1 326 of 2009 $ in years 2001, 2004, and 2008 respectively. This represents 2.2% annual decline in server costs between 2001 and 2004. If we complement this outcome with Employment Cost Index (ECI) obtained from FRED (2016), we find out that the index for Q4 2001 was 90.0 while in Q4 2004 was 97.7. This represents 2.8% yearly increase in cost. If we add weights to particular cost drivers according to proportion to total cumulative cost 3 months after startup, we obtain total annual cost change ΔC

$$\Delta C = 0.7\Delta SC + 0.3\Delta ECI \tag{3.1}$$

Trends in decrease in cost of IT equipment and rise in Employment Cost Index remained the same for the period 2004-2008. Therefore, we calculate average annual rate of decrease of costs associated with purchase and possession of IT equipment as -1% between years 2001 and 2005. From this, we arrive at startup costs of $115 865 in 2005.

3.2 AWS and comparison

Regarding AWS in 2006, the cost of small instance[2] was \$0,10 per hour, which represented cost of \$73 per month. Large instance[3] cost \$0.40 per hour which resulted in \$292 per month. We assumed that own IT solution was designed for 3-year period with sufficient room to scale and therefore was underutilized at least for the first year. On AWS, entrepreneurs are flexible to scale and can use only currently demanded bandwidth. Therefore, we calculate AWS cost for the first year on small instance server assuming that customers will upgrade to large instance server for years 2 and 3. To obtain results comparable to TCO of owned IT infrastructure, we do not include cost of downtime to calculation of total AWS cost estimation.

Our comparison yields results which are consistent with statements of recognized venture capitalists and entrepreneurs. With utilization of AWS, upfront costs in 2006 were reduced by approximately \$80 362 and cumulative expenditures after 3 months of operation were lowered by \$115 646 which is 529 fold decrease in relative terms. (see Table 3.1) We are aware of strong assumptions imposed on our estimations and variation in startups' IT requirements. We also realize that we neglected other factors (such as development of average amount of data transferred per user) that could influence server requirements and subsequently its cost. However, we presented conservative results for hardware configuration which would be sufficient to run basic applications and withstand several thousands of visits per day. With increasingly demanding IT infrastructure, the upfront costs will, not surprisingly, represent higher portion of TCO. (TechWise 2004) This will result in even higher savings in companies' early days when using AWS.

Table 3.1: Comparison of expenditures on computing resources (in \$)

	1 month	3 months	6 months	12 months
HP AlphaServer cluster	94 090	115 865	153 688	225 207
AWS	73	219	438	876
Change	1 289x	529x	351x	257x

[2] 1.7 GB of memory, 1 virtual core with 1 EC2 Compute Unit, 160 GB of instance storage
[3] 7.5 GB of memory, 2 virtual cores with 2 EC2 Compute Units each, 850 GB of instance storage

Chapter 4

Methodology and Data

In this chapter, we examine the impact of decreased barriers to entry on VC investment activity. We use Autoregressive Distributed Lag (ARDL) model which enables us to analyze relationships among our variables. This model is particularly appealing because of possibility to consider lags of dependent and independent variables and to separate long-run and short-run effects. In this chapter, we test 3 hypotheses

1. a negative shock in startup costs positively influences seed investments,
2. macroeconomic factors do not influence seed investments,
3. a shock in cost of starting a company does not influence later-stage investments.

In order to test our hypotheses, we will construct four models: two models with seed investments as a dependent variable and two models with investments in later stages as a dependent variable.

Regarding dataset, we include NVCA investment data and complement it with macroeconomic and selected technological factors. We chose macroeconomic variables based on research of Gompers & Lerner (1999), Jeng & Wells (2000) and Prohorovs & Pavlyuk (2013). Variables that are most significant in determining VC activity are portion of research & development on gross domestic product, interest rate, market indexes and government subsidies. As a market index we chose Nasdaq which contains stocks of major tech and Internet-related companies[1] and is therefore most likely to suit our purposes.

4.1 Methodology

In order to arrive at generall ARDL, we have to begin with dynamic regression model. It is a model which contains lagged explanatory variables to account for the time adjustment process. We express it by equation

$$y_t = \alpha + \sum_{i=0}^{r} \beta_i x_{t-i} + \epsilon_t \tag{4.1}$$

where y_t and x_t are random variables, α is constant, and ϵ_t denotes error term.

Here one-time change in independent variable at any time will impact $E[y_t|x_t, x_{t-1}, ...]$ in every following period. The duration of this change can last for either infinitely long (infinite lag models) or for a limited period of time (finite lag models).

For convenience, we will assume lag operator defined as

$$Lx_t = x_{t-1}. \tag{4.2}$$

This function has a property of lagging variables according to its power

$$L(Lx_t) = L^2 x_t = x_{t-2}. \tag{4.3}$$

Further, for clarity, we will define first difference as

$$\Delta x_t = x_t - x_{t-1}. \tag{4.4}$$

Thus, we can rewrite general dynamic model in a polynomial form

[1] Companies such as Google, Apple, Microsoft or Facebook are listed on Nasdaq

$$y_t = \alpha + \sum_{i=0}^{r} \beta_i L^i x_t + \epsilon_t = \alpha + B(L)x_t + \epsilon_t, \quad (4.5)$$

where α is constant and ϵ_t is disturbance.

Further, $B(L)$ represents polynomial in L and can be expressed by equation

$$B(L) = \beta_0 + \beta_1 L + \beta_2 L^2 + ... + \beta_r L^r \ . \quad (4.6)$$

By adding lags of dependent variable into equation 4.5, we arrive at special case of dynamic model called ARDL represented by equation

$$y_t = \mu + \sum_{i=1}^{p} \gamma_i y_{t-i} + \sum_{j=0}^{r} \beta_j x_{t-j} + \delta w_t + \epsilon_t \quad (4.7)$$

where μ is constant and $\delta w_t, \epsilon_t$ are error terms.

We can rewrite this equation in polynomial form

$$C(L)y_t = \mu + B(L)x_t + \delta w_t + \epsilon_t \quad (4.8)$$

by defining polynomial in the lag operator as

$$C(L) = 1 - \gamma_1 L - \gamma_2 L^2 - ... - \gamma_p L^p. \quad (4.9)$$

For convenience we define aggregated error term

$$\xi_t = \delta w_t + \epsilon_t. \quad (4.10)$$

We then arrive at equation

$$C(L)y_t = \mu + B(L)x_t + \xi_t, \quad (4.11)$$

which is equivalent to

$$y_t = \mu + \sum_{i=1}^{p} \gamma_i y_{t-i} + \sum_{j=0}^{r} \beta_j x_{t-j} + \xi_t \qquad (4.12)$$

where we assume ξ_t to be serially uncorelated.

This form can be denoted as $ARDL(p, r)$ where p and r indicate orders of polynomials $C(L)$ and $B(L)$ respectively. This model can be modified to numerous special cases by imposing restrictions on p and r. (Hendry *et al.* 1982)

In order to check for stability of our model, we need to ensure that autoregressive process is stationary. Therefore, the autocorrelation coefficient ρ have to fulfill condition $|\rho| < 1$. This can be tested for example by Augumented Dickey-Fuller (ADF) test or KPSS test. Greene (2003) also proposes possible problem that can arise from polynomial $C(L)$ equaling 0. When we reformulate polynomial form of ARDL to distributed lag form (4.13), we can immediately see the problem with term $\mu/C(L)$. It means that if $\sum_{i=1}^{p} \gamma_i$ =1 the model is unstable.

In determination of appropriate lag length in ARDL(p,r) model, Greene (2003) suggests to add lags until t-test shows statistical insignificance in the last-added lag. However, this approach suffers from omitted variable problem when we include less than p lags and therefore the estimator might be biased and inconsistent. This problem can be eliminated by starting with model with p lags included. Although p is unknown, we can specify value larger than p and regress y on p+d lags which yields consistent estimates of coefficients. Arriving to optimal lag lengths is done through minimization of Akaike information criterion (AIC) or Bayiesian information criterion (BIC).

To estimate coefficients in ARDL model, we can use Ordinary Least Squares (OLS) if it meets Gauss-Markov assumptions. Under these conditions, OLS is the best unbiased linear estimator, testing procedures are asymptotically valid and we can use F-statistics to perform tests. If it also fulfills an assumption of normality of residuals it is the best unbiased efficient estimator.

To interpret ARDL model, we modify the polynomial form to distributed lag form by division of the original equation by C(L). We obtain following equation

$$\begin{aligned} y_t = \frac{\mu}{C(L)} + \frac{B(L)}{C(L)} x_t + \frac{1}{C(L)} \xi_t = \\ = \frac{\mu}{1 - \gamma_1 - \gamma_2 - ... - \gamma_p} + \sum_{j=0}^{\infty} \alpha_j x_{t-j} + \sum_{l=0}^{\infty} \theta_t \xi_{t-l} \end{aligned} \tag{4.13}$$

Coefficients on lagged variable x are the individual terms in polynomial $\frac{B(L)}{C(L)}$. Lets denote them α_0, α_1, α_0,... for $1, L, L^2$ lagged variables.

Given

$$C(L)A(L) = B(L) \tag{4.14}$$

where

$$A(L) = (\alpha_0 L^0 - \alpha_1 L - \alpha_2 L^2 - ...). \tag{4.15}$$

In order to determine long-run effect of a change in independent variable, we can specify the equilibrium multiplier (Long-Run Multiplier)

$$LRM = \sum_{i=0}^{\infty} \alpha_i = A(L) = \frac{B(L)}{C(L)} = \frac{\sum_{i=0}^{r} \beta_i}{1 - \sum_{i=0}^{p} \gamma_i}. \tag{4.16}$$

Now we can find solution for each coefficient by solving set of linear equations for each α in

$$
\begin{aligned}
L^0 &: \alpha_0 = \beta_0 \\
L^1 &: \alpha_1 - \gamma_1\alpha_0 = \beta_1 \\
L^2 &: \alpha_2 - \gamma_1\alpha_1 - \gamma_2\alpha_0 = \beta_3 \\
&... \\
L^r &: \alpha_r - \gamma_1\alpha_{r-1} - \gamma_2\alpha_{r-2} - ... - \gamma_r\alpha_0 = \beta_r \\
L^{r+1} &: \alpha_{r+1} - \gamma_1\alpha_r - \gamma_2\alpha_{r-1} - ... - \gamma_{r+1}\alpha_0 = 0 \\
&... \\
L^p &: \alpha_p - \gamma_1\alpha_{p-1} - \gamma_2\alpha_{p-2} - ... - \gamma_p\alpha_0 = 0
\end{aligned} \tag{4.17}
$$

In recent literature, controlling for cointegration in ARDL model received great amount of attention. However, Bentzen & Engsted (2001) suggests that treating model with cointegration by error correction model or similar approaches does not yield significant improvement. We will check for cointegration in our model using Engle-Granger test.

4.2 Data

4.2.1 VC investment data

In this thesis, we use quarterly data published in PricewaterhouseCoopers (2016) report conducted in collaboration of NVCA and PricewaterhouseCoopers (PwC) who aim to deliver official comprehensive dataset on VC activity in US. We are concerned primarily with investment data which are obtained from numerous sources. Data is gained from surveys carried out by PwC's MoneyTree Report and NVCA based on Thomson Reuters data. It is complemented by SEC filings and publicly accessible resources and publications. The original dataset, from which majority of information is derived, is Thomson Reuter's ThomsonOne.com VentureXpert database of private equity and VC deals which contains 128 000 portfolio companies and 21 000 private equity firms.[2]

MoneyTree Report records investment activity of professional entities focused on financial investing in companies located in the US. These investors include,

[2] As of January 2016

among others, VC firms, corporate venture investing institutions or investment banks. Angel investments are recorded only when they are part of a financing round led by some of above mentioned investing institutions. This report excludes debt financing, Initial Public Offerings (IPO), recapitalization, buyout investments and non-cash private equity. It also excludes direct corporate investments unless they are syndicated investments, or they are proven to be financial investments, not outsourcing. Incubator investments are not recorded as long as they do not participate in syndicate investment with VC firms or other financial institutions that meet the conditions. Importantly, this dataset takes into account investments when cash was already exchanged for equity. Traditional methodology recognizes investment in time of signing term sheet. Therefore, the investment volume might be slightly different from information sources on VC investments which use traditional methodology.

Our dataset contains quarterly VC investment data from Q1 1995 till Q4 2015. As we mentioned above, VC industry was very unstable around year 2000 what would significantly distort our results. Hence, we ommit data affected by the Internet bubble and reduced our observations to a sample starting Q1 2001 and ending Q4 2015.[3] The sample contains 60 observations and summarizes total invested amount and total number of deals broken down by investment stage. Stages are divided into 4 groups; seed, early, expansion and later. We also possess data on total Internet-related VC investments which allows us to better examine the impact of AWS on VC activity by a more representative sample. According to NVCA, Internet-related companies are those having their primary technology application in a category defined as Internet-specific. To describe it on an example, Uber's primary business is transportation, however, its technology application is Internet-related and therefore, when reporting investment in Uber, it is classified as an Internet-related investment. Unfortunately, we do not possess data on Internet-related investments broken down by stage. It does not present a real problem in our research as Internet-related investments represent majority of total investments. Also, the classification of Internet-related companies is not complete and might be imprecise. Therefore, performing analysis on total investments yields consistent results in case of estimating impact of AWS on seed investment activity. This issue is further examined in chapter 5.

[3] To see more detailed explanation, see section 5.1

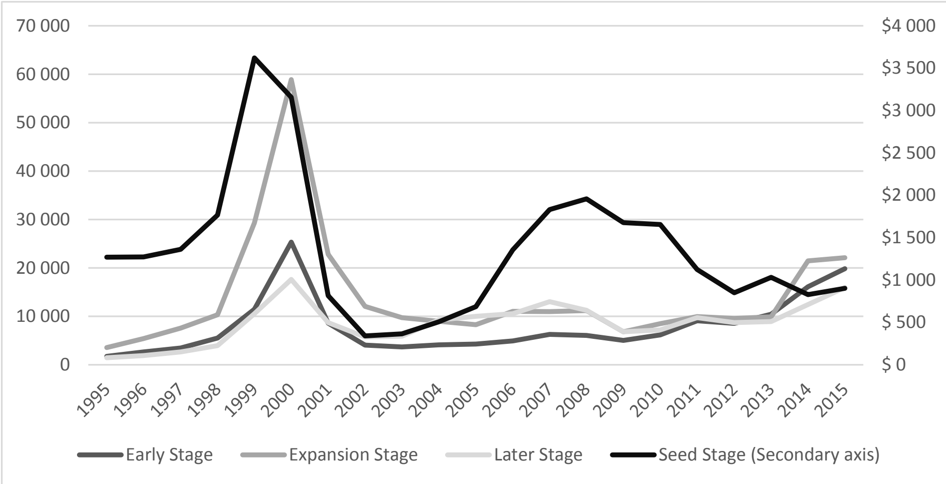

Figure 4.1: Total VC investments by stage (in $ Millions)

4.2.2 Macroeconomic factors

In our analysis, we include quarterly data of 4 macroeconomic indicators. The first is government investments into research and development (R&D) in US. The second is Gross Domestic Product (GDP) of the United States which can be seen as a demand-side effect on VC investment activity. Both indicators are seasonally adjusted and collected from FRED (2016). Further, we collected data on NASDAQ index from Yahoo (2016). The last indicator is US 10Y government bond yield, which is a representative of US interest rate. This variable was also collected from FRED (2016) and can be considered as a supply-side indicator of VC investments.

4.2.3 Factors potentially influencing startup costs

We also take into account three other factors that are potent to decrease cost of starting and operating businesses. These include open-source software, seed accelerators and business incubators, and MySQL technology. To estimate usage of these factors, we consider Google search volume for above mentioned terms. Google Trends tool provides this data on a weekly basis for period 1.1.2004 - 31.12.2015. Hence, data on this group of factors covers shorter period and will be analyzed separately. Google Trends allows to analyze search volume from Google over a given period of time in a certain location and scales results to a range from 0 to 100. Index of 0 corresponds to the relatively lowest search volume and 100 to the highest in a given time and place. We used data starting 1.1.2004 in the United States.

4.2.4 Barriers to entrepreneurship and government subsidies

The last group of factors that might have influenced seed investments in 2006 is related to government spending and ease of starting a business. We studied data on government financial assistance for small businesses in USA by examining a number of official information sources.[4] As an indicator of barriers to entrepreneurship, we use ease of starting a business index[5] conducted by World Bank Group.

4.3 Empirical model

In this section, we construct variables and examine their properties. Then, we adjust them accordingly and proceed with specification of four models using Autoregressive Distributed Lag (ARDL) approach. We also present results on consistency of our models.

4.3.1 Construction of variables

In our models, we use variables related to VC investment activity and variable representing cost of starting an Internet-related startup. The first group of variables contains VC investments broken down by stage of investment. One variable represents seed investments (Inv_SEED) and the second represents sum of early, expansion and later stage investments (Inv_LATER). Hereinafter we refer to this sum as later-stage investments. Variable representing cost of starting a company is constructed from values computed in Chapter chapter 3. We estimated cost in 4 time points; in 2001, 2005, 2006 and 2015. We then estimated cost between these points by simple linear interpolation and denoted it $COST$. In accordance with launch of EC2, costs dropped in Q2 2006.

[4] https://www.usaspending.gov, https://www.sbir.gov, https://www.sec.gov, https://www.sba.gov

[5] http://www.doingbusiness.org

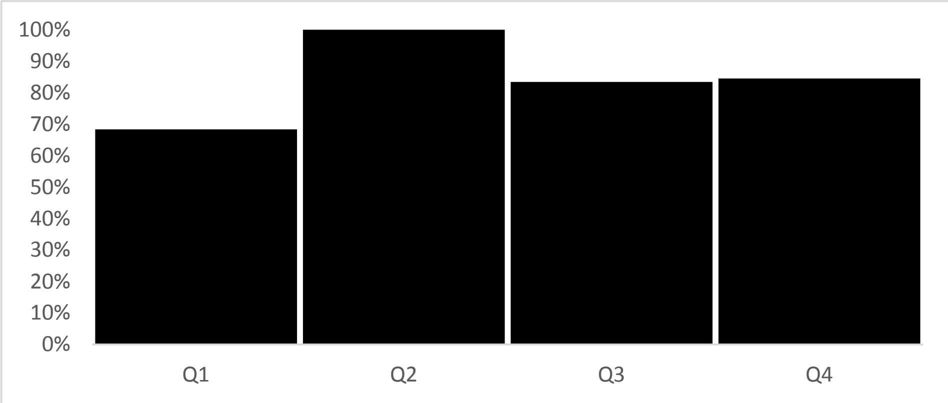

Figure 4.2: Total seed investments by quarter

Regarding macroeconomic factors, we constructed variables NASDAQ, RDGDP and IR for Nasdaq index, ratio of R&D and GDP and US interest rate, respectively.

Factors that might influence cost of startup are constructed by grouping indexes for related keywords. Indexes are then averaged to obtain single index which represents search volume for a particular factor. We constructed indexes related to MySQL, open-source, and seed accelerators and business incubators.

Adjusting for seasonality

By examining periodicity in our data, we find that both investment variables contain high levels of seasonality (See Figures 4.2 and 4.3). Therefore, we seasonally adjust those variables by using TRAMO/SEATS [6] method developed by Victor Gomez and Agustin Maravall. It is an approach recommended by ESS Guidelines on Seasonal Adjustment[7] and is broadly used by major statistical institutions. For easier interpretation, we logarithmized both investment variables. We denote transformed variables $l\_Inv\_SEED\_sa$ and $l\_Inv\_LATER\_sa$.

[6]Time Series Regression with ARIMA Noise, Missing Observations, and Outliers/ Signal Extraction in ARIMA Time Series

[7]http://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-15-001

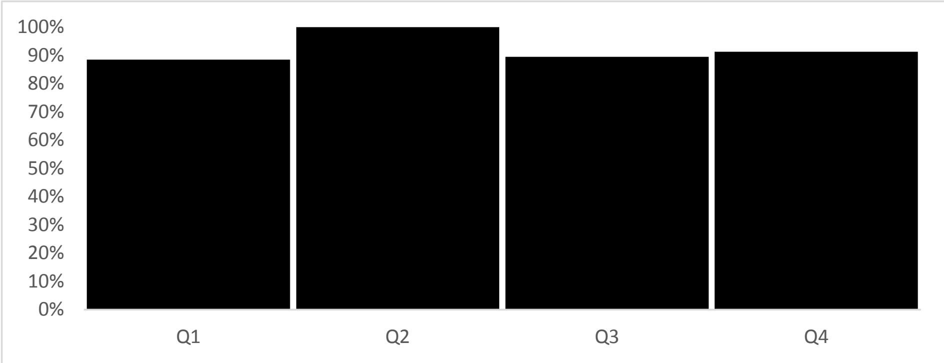

Figure 4.3: Total later-stage investments by quarter

Unit root test

In our analysis we test covariance-stationarity which is defined as stochastic process with finite and constant variance, constant mean, and $Cov(x_t, x_{t+h})$ dependent only on t, not h. Generally, most of time-series data are non-stationary. It presents problems such as spurious regression and causes standard assumptions to be asymptotically invalid. There are two types of non-stationarity in time-series: deterministic and stochastic. Both cases require different treatments. Time series with deterministic trend can be transformed to stationary by substraction of the trend while stochastic non-staionarity can be treated by first differencing.

To test the stationarity of our variables, we used Augumented Dickey-Fuller (ADF) and KPSS tests. ADF test has the null hypothesis of unit root while KPSS's null hypothesis is that data is stationary. We present results of both tests in table 4.1. Both methods have certain drawbacks and it is common that tests show contradictory results. ADF tend to be imprecise on small samples and underperforms KPSS under such conditions. Therefore, considering our sample of only 60 observations, we will prioritize KPSS test where results are equivocal.

Results of unit-root tests show contradiction present only in $l\_Inv\_SEED\_sa$. Here, according to ADF test, we can reject null hypothesis of unit root with p-value = 0.003 while KPSS shows that we can reject hypothesis of no unit root with p-value = 0.044. By checking autocorrelation function (ACF) of this variable, we can observe high number of slowly decreasing significant spikes. This is an indicator that series has a long memory and are therefore non stationary.

The rest of variables are non stationary according to both tests.

Table 4.1: Unit-root tests

	ADF		KPSS	
	t-statistic	p-value	t-statistic	p-value
$l\_Inv\_SEED\_sa$	-3.82866	0.003*	0.499727	0.044*
$l\_Inv\_LATER\_sa$	-0.13009	0.218	0.530636	0.040*
$COST$	-1.42681	0.563	0.600336	0.029*
$NASDAQ$	1.34974	0.999	0.616144	0.027*
IR	-1.0547	0.736	0.648901	0.022*
$RDGDP$	-1.88511	0.340	0.469548	0.049*

*The null hypothesis can be rejected at 5% level.

By checking plots of our variables we choose to control for stationarity by first differencing. We repeat ADF and KPSS tests on first-differenced variables (See Table 4.2). Both tests suggest stationarity of all variables except $d\_l\_Inv\_SEED\_sa$, which is non-stationary according to ADF. As we mentioned above, KPSS is more suitable for small samples and our variable is stationary according to this test. We can conclude that all variables are integrated of order 1, denoted as I(1).

Table 4.2: Unit-root tests of first-differenced variables

	ADF		KPSS	
	t-statistic	p-value	t-statistic	p-value
$d\_l\_Inv\_SEED\_sa$	-3.1008	0.106	0.138731	> 0.10
$d\_l\_Inv\_LATER\_sa$	-7.14034	< 0.01*	0.31713	> 0.10
d_COST	-7.674	< 0.01*	0.13624	> 0.10
d_NASDAQ	-5.8896	< 0.01*	0.07119	> 0.10
d_IR	-6.74331	< 0.01*	0.07356	> 0.10
d_RDGDP	-3.54376	0.044*	0.060584	> 0.10

*The null hypothesis can be rejected at 5% level.

Cointegration

Another problem that might arise in our model is cointegration. It is defined for a set of I(1) time-series variables. If a linear combination of this set exists and it is I(0), then we can say that variables are cointegrated. This might lead to a problem of spurious regression.

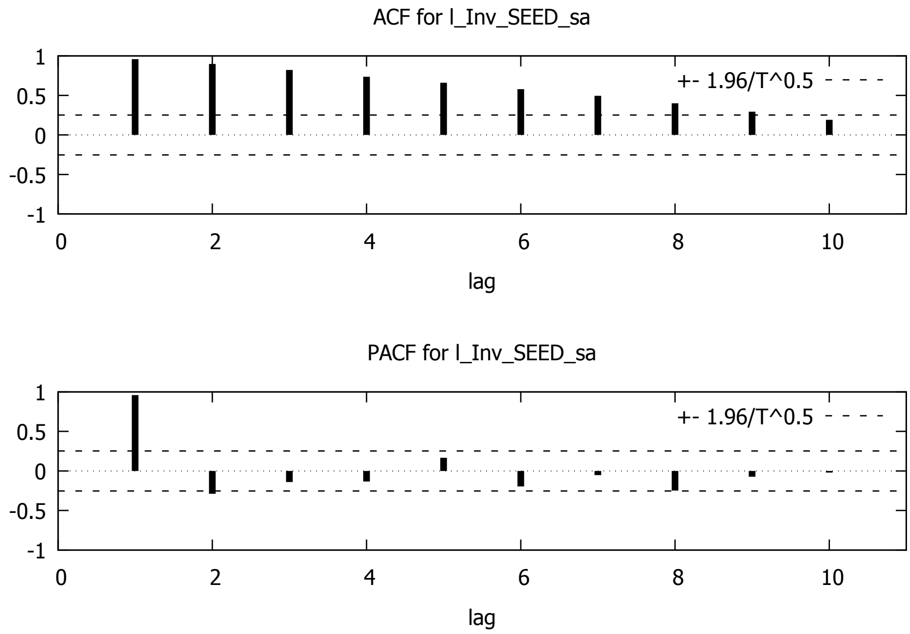

Figure 4.4: ACF and PACF for *l_Inv_SEED_sa*

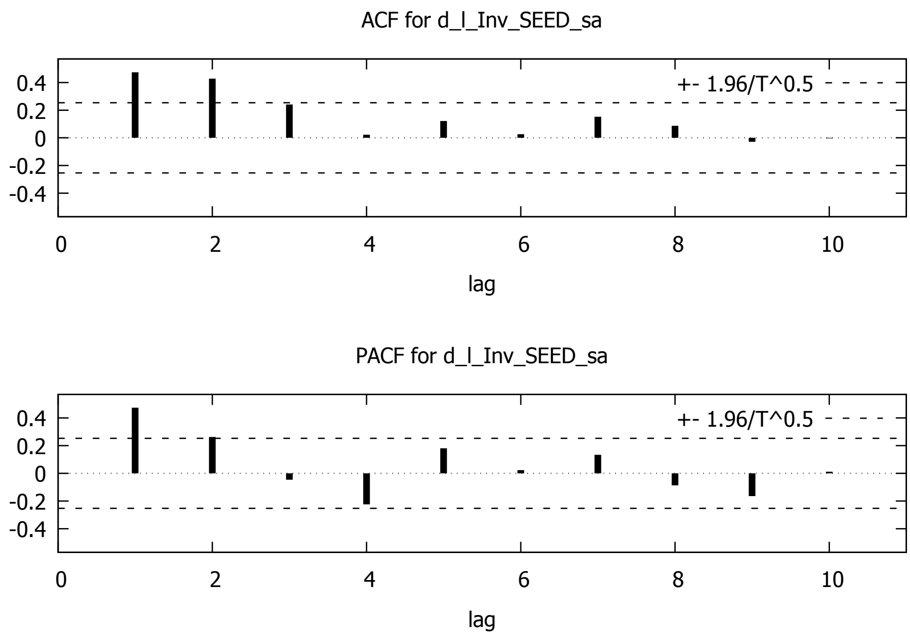

Figure 4.5: ACF and PACF for *d_l_Inv_SEED_sa*

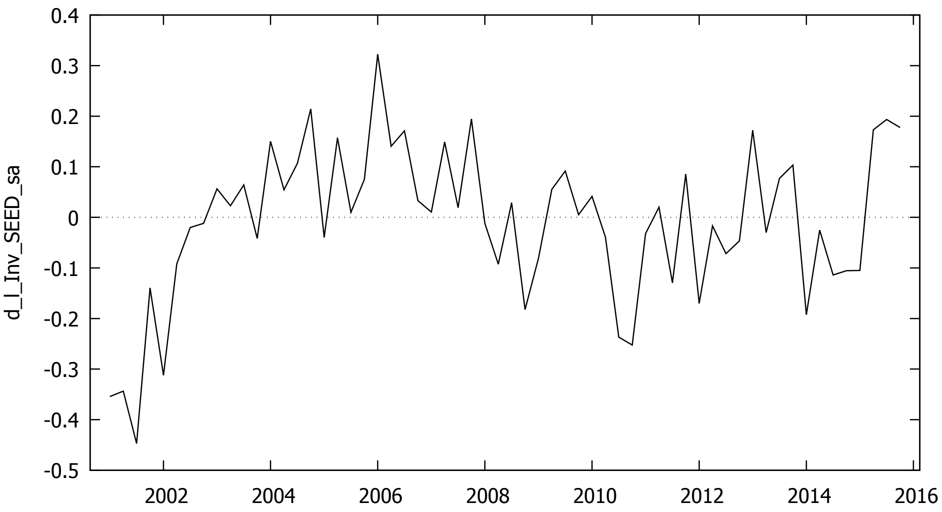

Figure 4.6: Plot of $d\_l\_Inv\_SEED\_sa$

Our variables are I(1) and hence fulfill the first assumption of cointegration. We test for presence of this property using Engle-Granger test on set of 5 variables $l\_Inv\_SEED\_sa$, $COST$, $NASDAQ$, IR, $RDGDP$ and set of 2 variables $l\_Inv\_LATER\_sa$, $COST$. In both sets, the residuals from auxiliary regression are not I(0), thus we can reject cointegration.

4.3.2 Model specification

In order to test our hypotheses, we modify general form of ARDL model represented by equation (4.7). The first hypothesis is whether decrease in cost of starting a new company positively influence seed investments. We therefore estimate simple ARDL with dependent variable representing seed investments and independent variable standing for cost of starting a company with appropriate lag lengths.

In previous subsection we constructed variables which are adjusted for seasonality and we ensured that they are also stationary. In order to determine the optimal lag length in our model we minimize AIC and BIC arriving at ARDL(2,0).

Model 1 can be rewritten as

$$\begin{aligned} d\_l\_Inv\_SEED\_sa_t = \mu + d\_l\_Inv\_SEED\_sa_{t-1} + \\ + d\_l\_Inv\_SEED\_sa_{t-2} + d\_COST_t + \xi_t, \end{aligned} \tag{4.18}$$

where μ is constant term and ξ_t represents error.

In order to determine whether OLS is a suitable estimator, we have to check Gauss-Markov assumptions. Estimated equation has linear form and $E[\xi_t|X] = 0$. Also, the sample is random and variables do not show signs of collinearity. This means that OLS is an unbiased estimator. White's and Breusch-Pagan tests for heteroskedasticity show homoskedastic residuals. Moreover, residuals have normal distribution which makes OLS best unbiased estimator and is therefore suitable for our estimation.

Model 2 is an extension of the first regression and is supposed to test the second hypothesis. It states that other factors did not influenced seed investments. In this model, we add 3 first-differenced macroeconomic variables and cost of startup as regressors (d_RDGDP, d_IR, d_NASDAQ, d_COST). The model is estimated using modified ARDL with 4 explanatory variables. We will denote it $ARDL(p, q_1, q_2, q_3, q_4)$ which is represented by general equation

$$y_t = \mu + \sum_{i=1}^{p} \gamma_i y_{t-i} + \sum_{j=0}^{q_1} \beta_j x_{t-j} + \sum_{k=0}^{q_2} \delta_k x_{t-k} + \sum_{l=0}^{q_3} \epsilon_l x_{t-l} + \sum_{m=0}^{q_4} \theta_m x_{t-m} + z_t, \tag{4.19}$$

where μ is constant and z_t error term.

Examining correlogram of each variable, we determine lag lengths for the 3 macroeconomic regressors. d_RDGDP's PACF has only one significant spike at lag 1 meaning that any higher-order autocorrelation is explained by the first lag of the variable. d_IR's PACF has significant spike at lag 2 suggesting to add 2 lags in ARDL. Finally, d_NASDAQ has no significant spikes in its PACF and therefore we do not include any lags of this variable. d_COST and $d\_l\_Inv\_SEED\_sa$ have 0 and 2 lags respectively, as we revealed in the Model 1. We, hence, arrive at Model 2

$$
\begin{aligned}
d\_l\_Inv\_SEED\_sa_t = {} & \mu + d\_l\_Inv\_SEED\_sa_{t-1} + d\_l\_Inv\_SEED\_sa_{t-2} + \\
& + d\_COST + d\_RDGDP_t + d\_RDGDP_{t-1} + \\
& + d\_IR_t + d\_IR_{t-1} + d\_IR_{t-2} + d\_NASDAQ + z_t.
\end{aligned} \tag{4.20}
$$

By checking Gauss-Markov assumptions, we find out that the model fulfils the conditions and thus OLS is an efficient estimator.

Moving to Model 3, our last hypothesis states that investments in later stages are not significantly influenced by shock in cost of starting a company. This involves regression of first-differenced logarithm of investments in later stages on first difference of costs with lags of both dependent and independent variables. By repeating the above described procedures, we arrive at model with zero lags in both regressor and regressand. This is a special case of ARDL which is identical with OLS.

$$
d\_l\_Inv\_LATER\_sa_t = \mu + d\_COST_t + \epsilon_t \tag{4.21}
$$

The third model fulfils all six Gauss-Markov assumptions and OLS is therefore an efficient estimator.

We also estimate Model 4 with $d\_l\_Inv\_LATER\_sa$ as dependent variable and current and lagged values of $COST$ and macroeconomic variables.

$$
\begin{aligned}
d\_l\_Inv\_LATER\_sa_t = {} & \mu + d\_COST + d\_RDGDP_t + d\_RDGDP_{t-1} + \\
& + d\_IR_t + d\_IR_{t-1} + d\_IR_{t-2} + d\_NASDAQ + z_t,
\end{aligned} \tag{4.22}
$$

where OLS is an efficient estimator as well.

Chapter 5

Results

In this chapter, we describe development of considered variables and interpret the models. Model 1 (4.18) and Model 2 (4.20) verify hypotheses 1 and 2 respectively. Model 3 (4.21) and 4 (4.22) are concerned with third hypothesis. We interpret effect of startup costs on dependent variables, compare models and make conclusions in broader economic context.

5.1 Descriptive statistics

5.1.1 Early years

The VC industry encountered relatively unstable period in the past two decades. The most striking event in our dataset is the Internet bubble, which culminated in 2000 and which is responsible for huge leap in both, total invested amount and number of deals in VC industry across all stages. VC activity rocketed from $21 562M in 1998 to a high of $104 998M in 2000 and number of deals leapt from 3 744 in 1998 to 8 041 deals in 2000. This shock is particularly tangible in Internet-related investments which accounted for roughly 77% of total capital invested during the Internet bubble with average investment size of $17.48M. After reaching sky-high values, investment activity plummeted to $40 939M the year later and continued to sink until 2003 stopping at a low of $19 682M.

If we take a closer look at more granular data, showing investment activity broken down by stage, increase in seed investments was not as steep as overall VC investment activity. Seed registered substantial upsurge in 1999 rising from $1 767M to its climax of $3 621M. Here, the decline showed the year earlier compared to total investment which reached its highest value in 2000. The biggest drop was in 2001 when seed plummeted by 74% and slipped even deeper the year after to its low of $341M.

5.1.2 Era of modern VC

From 2003, total VC investments maintained relatively low volatility with steady growth up until global financial crisis in 2008. Over this period, VC activity registered average yearly growth of 7.5%. VC investment activity was hit by financial crisis in full strength in 2009 falling by 33.18%. During this period, companies received record low average investments amounting only $6.41M which represents 63% decline from 2000. From here, VC industry climbed rapidly with average 19.4% yearly growth until 2015 with only minor bump in 2012. Investors were gradually gaining confidence which is reflected in increasing average deal size toping $13M in 2015.

Regarding seed investment, we can see constantly increasing trend over the period 2002 - 2005. In 2006, we can observe substantial jump accounting for 97.6% increase to a high of $1 357M. Comparing it to sum of investments in later stages, where we can observe only 17.5% increase, evidence suggests that seed investments encountered significant shock. Seed maintained rapid growth even the year after with 35% increase while later stages together grew only by 14.2%. Growth of seed remained highly above later stages until 2010, when it underperformed later stages for the first time since 2006. After 2010, seed investments plummeted to $851M in 2012 which represents 48.6% drop. Since then, seed investments have been rather stagnant while later stages have been on the rise.

Internet bubble is a market anomaly which caused extreme deflection in investment activity. It skews our analysis and in order to present accurate results we omit it from our further analysis. In our model we examine sample starting Q1 2001 and ending Q4 2015.

5.1.3 Macroeconomic and other factors

To start with macroeconomic factors, we examine NASDAQ, IR and RDGDP. NASDAQ has increasing overall trend with only two bumps corresponding to burst of the Internet bubble and the global financial crisis. On the other hand, IR has declined since 2001 from 5% to roughly 2% while RDGDP remained constant at slightly below 1%. Focusing on 2006, NASDAQ and RDGDP remained steady while IR jumped in the second quarter from 4.57% to 5.07% falling back to 4.63% in the fourth quarter. As increased interest rates discourage additional investment spending, we cannot say that IR is correlated with increase in seed investments. In this period, we can observe negligible jump and therefore cannot say that IR is correlated with seed investments.

Regarding other factors that are potent to influence barriers to entry, we can observe overall declining trend in all three variables (See Figure 5.1). In 2006, Accelerators & Incubators and MySQL were declining constantly without any fluctuation. Open-source was more volatile with its jump from 59 to 65 in Q1 2006 accounting for 9% increase, however, it slipped back in the following quarter.

In the light of above mentioned circumstances, there is a support in favor of our hypothesis that emergence of AWS positively influenced seed investments through decreased barriers to entry. However, in order to conclude more certainly, we will further examine our data in the following section.

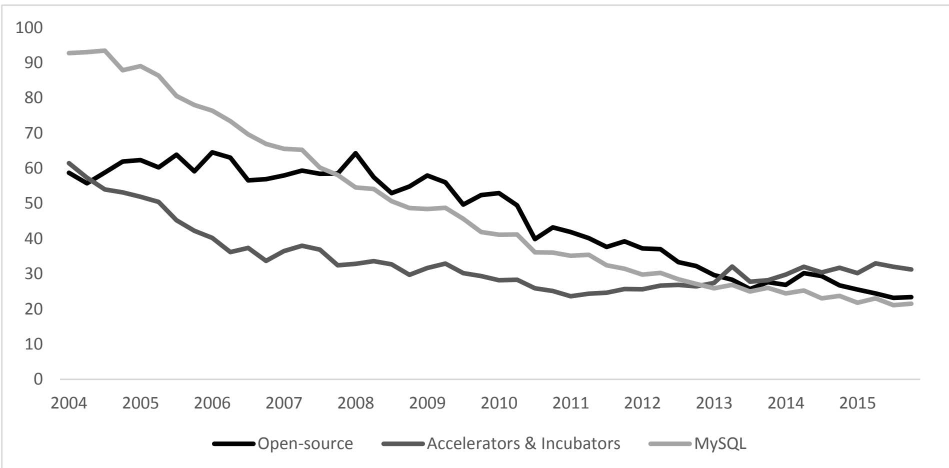

Figure 5.1: Interest in selected technological factors

5.1.4 Summary statistics

In order to examine our sample more thoroughly, we present summary statistics (5.1) for selected variables. Our analysis shows that all variables are slightly positively skewed except $COST$ which is skewed substantially positively. This is result of distribution of values with long right tail. Regarding kurtosis, all variables except Inv_LATER show negative kurtosis resulting in flatter distribution. $COST$ is again far from normal distribution with kurtosis -1.50 while Inv_SEED, No_SEED and No_LATER are flattened only slightly. Inv_LATER is heavily spiked with value of 2.33. As of dispersion, all variables except Inv_LATER show relatively low spread with coefficient of variation (CV) less than 1. Inv_LATER has coefficient of 1.66 meaning its distribution is dispersed relatively more.

Table 5.1: Summary statistics

	Mean	Median	Min	Max	SD	CV	Skewness	Kurtosis
Inv_SEED	2.65e+08	2.34e+08	7.64e+07	6.77e+08	1.48e+08	0.56	0.76	-0.17
Inv_LATER	7.41e+09	6.55e+09	3.53e+09	1.71e+10	2.99e+09	0.40	1.66	2.33
$COST$	3.97e+04	207	65	1.29e+05	5.63e+04	1.42	0.71	-1.50

5.2 Impact of AWS on seed and later-stage investments

Our results from Model 1 show significant negative correlation between seed investments and cost of starting a company (See table 5.2). Interpreting results according to set of equations 4.17 in lag 0 yields $\alpha_0 = \beta_0$. This model is in log-level form and hence one-unit change in d_COST results in $100 * \alpha_0$ % change in $d\_l\_Inv\_SEED\_sa$. With coefficient α_0 = -2.566e-06 and p-value of 0.013, we can say that the correlation is significant and with \$1 000 decrease in cost of starting a company, there will be approximately 0.26% increase in seed investments. If we relate this result to the introduction of AWS when costs dropped by \$115 646, we can observe 29.67% increase in seed investments. In this case we interpret only short-run effect as we have no lags in variable d_COST. In Section 5.1 we see seed investments almost double in 2006. These results suggest to confirm the first hypothesis that seed investments are influenced by cost of starting a company. However, we will test also Model 2 in order to find other factors that might be originators of the change and compare both models.

The Model 2 also showed d_COST to be statistically significant with p-value = 0.010 (See table 5.2). The variable is not lagged and therefore, coefficient $\beta_0 = \alpha_0$ as well. Coefficient α_0 is comparable to the previous model with value of -2.468e-06 which represents roughly 0.25% increase in seed investment when cost is reduced by $1 000. Other variables are statistically insignificant except the first lag of d_IR. To further examine impact of d_IR on seed financing, we estimate submodel of Model 2 which includes d_COST and current value of d_IR and its two lags as regressors. (See 5.1) We obtain coefficients that are not significant at 5% level on d_IR and both lags.

$$\begin{aligned} d\_l\_Inv\_SEED\_sa_t = \mu &+ d\_l\_Inv\_SEED\_sa_{t-1} + d\_l\_Inv\_SEED\_sa_{t-2} \\ &+ d\_COST_t + d\_IR_t + d\_IR_{t-1} + d\_IR_{t-2} + \xi_t \end{aligned} \tag{5.1}$$

Looking at coefficients of determination in Model 1, we can observe R-squared = 0.38 and adjusted R-squared = 0.34. This result shows model's relatively good description power. Model 2 has R-squared = 0.45 and constant adjusted R-squared = 0.34. This suggests that macroeconomic variables add negligible amount of new information to our model. Therefore, we chose Model 1 for interpretaion as it is more concise.

The first stability condition of Model 1 and 2 is met as we have stationary dependent and independent variables. We also have to ensure that sum of coefficients on lagged dependent variable is not equal to 1. We get $\gamma_{t-1} + \gamma_{t-2}$ = 0.58 which indicates that model is dynamically stable.

Regarding other factors that might have decreased barriers to entry and increased seed investments in 2006, MySQL have potential to reduce database TCO substantially as it is not required to pay for software licences. However, it was introduced in 90' and became standard in 00'. Also findings from Section 5.1 support our confidence that MySQL did not influence barriers to entry in 2006. MySQL could have impact on seed financing in 90' but to reveal it is beyond the scope of this thesis. Similarly, interest in business incubators and seed

Table 5.2: ARDL for seed investments

$d\_l\_Inv\_SEED\_sa_t$	Model 1	Model 2
$const$	1.236e-03 (1.516e-02)	3.628e-03 (1.599e-02)
$d\_l\_Inv\_SEED\_sa_{t-1}$	0.320* (0.109)	0.112 (0.118)
$d\_l\_Inv\_SEED\_sa_{t-2}$	0.340* (0.110)	0.345 (0.108)
d_COST_t	-2.542e-06* (1.002e-06)	-2.468e-06* (9.154e-07)
d_RDGDP_t	-	0.747 (2.924)
d_RDGDP_{t-1}	-	1.897 (2.567)
d_IR_t	-	0.035 (0.049)
d_IR_{t-1}	-	0.093* (0.042)
d_IR_{t-2}	-	0.004 (0.045)
d_NASDAQ	-	8.910e-05 (9.914e-05)

* Significant at 5% level

accelerators was experiencing stable decline around year 2006. The first seed accelerator, Y Combinator, which was founded in March 2005 invested only $200 000 which represents negligible fraction of total seed financing.[1] Open-source induced shift in economics of server TCO like Amazon did. However, this concept, similarly as MySQL, had its biggest impact in 90' while in 2006 it was perceived as industry standard. This is also supported by findings from Section 5.1.

In order to check for influence of government in our analysis, we examined public spending on small businesses and barriers to entrepreneurship in USA. We did not find any increased activity in 2005 and 2006. We proceed with examining barriers to entrepreneurship in USA. Index representing ease of starting a business in US remained constant over the period and did not show any deflection that could induce leap in seed investments. Based on above evidence, we confirm our second hypothesis which states that seed investments are not influenced by macroeconomic and technological factors.

By examining Models 3 and 4 we test our last hypothesis which states that later-stage investments are not influenced by cost of starting a business. Results from Model 3 showed no significant relationship between later-stage investments and cost of startup. Explanatory variable d_COST has p-value = 0.167 and is therefore insignificant. Moreover, this model has R-squared = 0.03 which further supports our hypothesis that change in cost does not influence later-stage investments. Model 4 shows relatively high R-squared = 0.30 and adjusted R-squared = 0.20. Nevertheless, coefficient at d_COST is still insignificant with even higher p-value = 0.180. The rest of regressors in Model 4 turned out to be also insignificant except d_NASDAQ which is significant at 5% level with coefficient equal to 3.023e-04. The significance of Nasdaq index in Model 4 and jump in value of coefficient of determination suggests that development of Nasdaq is positively related to later-stage investments.

Stability condition for Model 3 and 4 is met as all regressors and $d\_l\_Inv\_LATER\_sa$ are stationary and because these models lack autoregressive term, polynomial C has degree 0 and is therefore equal to 1. Model 3 suffers from ommited variable while Model 4 seems to be better fitted. Considerable increase in value of R-squared in Model 4 along with insignificance of d_COST in both models supports our last hypothesis that cost of starting a company does not significantly

[1] http://old.ycombinator.com/start.html

influence later-stage investments.

Regarding seed financing, our results are in line with our expectations. Companies in their earliest stages face extreme risks caused mainly by their instability and immaturity. Factors that directly affect operation of such company are substantially more determining for future success then macroeconomic changes. Macroeconomic factors also play their role in seed as they, for example, influence supply of capital but the effect is negligible relative to microeconomic factors. Seed involves experimentation and placing "bets" on numerous deals expecting exceptional returns only from few. Investments in such companies cannot be supported by previous performance as seed startups are usually pre-market. The idea, founding team and potential market are the only visible determinants of success. With launch of AWS, emerging companies were able to cut considerable part of their overall startup costs and gain comparative advantage. This was a great enticement for VC industry and it reflected in upsurge in total seed investments.

Later-stage companies are not directly affected by decrease in startup costs and effect of operational cost reduction is ambiguous among mature firms. Investing in more mature companies is perceived as not-so-risky compared to seed and is approached differently. On the other hand, later-stage investments are expected to have lower but more guaranteed returns. Investors have information on past performance and can make more founded decisions. Also, companies are more mature and less vulnerable to operational risks. Along with bigger size and allegiance to wide range of subjects, mature companies are more dependent on market sentiment and macroeconomic factors.

Table 5.3: ARDL for later-stage investments

$d\_l\_Inv\_LATER\_sa_t$	Model 3	Model 4
$const$	-0.011	-0.003
	(0.022)	(0.021)
d_COST_t	-2.026e-06	-1.602e-06
	(1.448e-06)	(1.187e-06)
d_RDGDP_t		4.031
	-	(3.787)
d_RDGDP_{t-1}		-5.638
	-	(3.317)
d_IR_t		0.063
	-	(0.063)
d_IR_{t-1}		0.060
	-	(0.054)
d_IR_{t-2}		0.024
	-	(0.056)
d_NASDAQ		3.024e-04*
	-	(1.281e-04)

* Significant at 5% level

Chapter 6

Conclusion

In the thesis we assess the impact of introduction of Cloud Computing on VC financing in the United States. In the first part of this thesis we evaluate the impact of Cloud Computing on cost of staring an Internet-related company. Amazon is considered a pioneer of modern Cloud Computing and made this technology available for general public. This consequently decreased cost of computing capabilities for Internet-related startups. To quantify this change we estimate total cost of ownership for entry-level IT infrastructure for period 2001-2005 based on a research of numerous reports. We complement the results with pricing and corresponding configurations from AWS for period 2006-2015 and construct cost development of computing capabilities for the whole period. For our purposes we consider costs associated with the first 3 months of company's operation. Results indicate substantial fall in cost of computing capabilities with use of AWS against owning IT infrastructure. In particular, the change accounts for approximately \$80 362 savings in upfront expenditures and a total of \$115 646 decrease in startup costs after 3 months of operation. In relative terms it means 529 fold decrease in startup costs in the 3-month time frame.

Analysis of our VC investment data, obtained from NVCA, shows that seed investments jumped by 97.6% in 2006 and maintained rapid growth for the two following years. On the other hand, later-stage investments remained relatively tranquil. This is partly a foundation for our hypotheses that negative shock in cost of starting a company positively influences seed investments but do not affect later-stage investments. To examine our hypotheses more thoroughly we use Autoregressive Distributed lag (ARDL) approach to construct 4 main models.

The main finding of this thesis is that introduction of AWS and subsequent startup cost reduction significantly negatively influences seed investments. According to results of Model 1, the drop in cost of starting a company induce 29.67% increase in seed investments. This finding suggests to confirm our first hypothesis. Further, we extend Model 1 and estimate the influence of Nasdaq market index, interest rate and portion of R&D on GDP on seed investments. Results show that the macroeconomic factors are not statistically significant and do not add significant amount of information to our model. This suggests to confirm the second hypothesis which states that macroeconomic factors do not influence seed investments. To examine factors, that are potent to influence seed investments, in more depth we analyze usage of MySQL, Open-source, and Business incubators and accelerators along with data on government subsidies and barriers to entrepreneurship. Findings show that all factors remained tranquil in the period when seed investments leapt. This suggests that seed investments were not influenced by these factors and that introduction of AWS remains the originator of the increase in seed.

The results are in line with our theory that decrease in costs associated with starting a company opened new entrepreneurial opportunities and induced influx of investments to new concepts. Also the insignificance of macroeconomic factors agrees with our supposition that companies in their earliest stages are more vulnerable to threats from microeconomic level than adverse changes in macroeconomic environment.

Further analysis shows statistical insignificance of impact of introduction of AWS on later-stage investments. This is also in line with our theory as more mature companies already own IT infrastructure and converting to AWS does not provide substantial savings. Moreover, the reduction in cost of computing capabilities is ambiguous in case of high-scale IT requirments.

This thesis aims to deepen the research of impact of Cloud Computing on Venture Capital. To our knowledge, this topic was researched only by Ewens *et al.* (2015) and is by no means exhausted. We believe that results from this field can be greatly enhanced by analyzing this problem on a more granular sample. Further research might aim at evaluation of investment performance of companies using Cloud against those owning IT infrastructure.

Bibliography

Bentzen, J. & T. Engsted (2001): "A revival of the autoregressive distributed lag model in estimating energy demand relationships." *Energy* **26(1)**: pp. 45–55.

Bernstein, S., A. G. Korteweg, & K. Laws (2015): "Attracting early stage investors: Evidence from a randomized field experiment." *Journal of Finance, Forthcoming* pp. 14–17.

Ewens, M., R. Nanda, & M. Rhodes-Kropf (2015): "Cost of experimentation and the evolution of venture capital." *Harvard Business School Entrepreneurial Management Working Paper* **(15-070)**.

EY (2015): "Venture Capital Insights - 4Q14." *Technical Report CY0832*, EYGM Limited.

FRED (2016): "Federal reserve economic data." `https://research.stlouisfed.org/fred2/`. Accessed: 20.2.2016.

Gompers, P. A. & J. Lerner (1999): "What drives venture capital fundraising?" *Technical report*, National bureau of economic research.

Gornall, W. & I. A. Strebulaev (2015): "The economic impact of venture capital: Evidence from public companies." *Available at SSRN 2681841* .

Greene, W. H. (2003): *Econometric analysis*, volume 5. Prentice hall.

Hendry, D. F., A. R. Pagan, J. D. Sargan *et al.* (1982): *Dynamic specification.* Social Science Research Council.

Jeng, L. A. & P. C. Wells (2000): "The determinants of venture capital funding: evidence across countries." *Journal of corporate Finance* **6(3)**: pp. 241–289.

KOOMEY, J. G., C. BELADY, M. PATTERSON, A. SANTOS, & K.-D. LANGE (2009): "Assessing trends over time in performance, costs, and energy use for servers." *Lawrence Berkeley National Laboratory, Stanford University, Microsoft Corporation, and Intel Corporation, Tech. Rep* .

KORHONEN, A. (2016): "How to find an angel and how to be an angel?" European Angel and P2P Summit 2016, Prague, Czech Republic.

NVCA (2016): "National venture capital association 2016 yearbook."

PAUL GOMPERS, J. L. (2001): "The venture capital revolution." *The Journal of Economic Perspectives* **15(2)**: pp. 145–168.

PRICEWATERHOUSECOOPERS (2016): "Moneytree report." `https://www.pwcmoneytree.com/`. Accessed: 20.2.2016.

PROHOROVS, A. & D. PAVLYUK (2013): "Analysis of economic factors influencing venture capital investment in european countries." *Social Research* **(4)**: p. 33.

SAMUEL KORTUM, J. L. (2000): "Assessing the contribution of venture capital to innovation." *The RAND Journal of Economics* **31(4)**: pp. 674–692.

TECHWISE (2004): "Total cost of ownership for entry-level and mid-range clusters." **(Version 1.0)**.

WILLIAMS, B. (2012): *The Economics of Cloud Computing.* Cisco Press.

YAHOO (2016): "Yahoo finance." `http://finance.yahoo.com/`. Accessed: 20.2.2016.

ZHANG, Q., L. CHENG, & R. BOUTABA (2010): "Cloud computing: state-of-the-art and research challenges." *Journal of internet services and applications* **1(1)**: pp. 7–18.

Appendix A

Appendix 1

Model 1: OLS, using observations 2001:3–2015:4 ($T = 58$)
Dependent variable: d_l_Inv_SEED_sa

	Coefficient	Std. Error	t-ratio	p-value
const	0.00401587	0.0152372	0.2636	0.7931
d_COST	−2.56593e–006	1.00023e–006	−2.5653	0.0131*
d_l_Inv_SEED_sa_1	0.249803	0.119197	2.0957	0.0408*
d_l_Inv_SEED_sa_2	0.333508	0.115362	2.8910	0.0055*

Mean dependent var	0.002362	S.D. dependent var	0.140729
Sum squared resid	0.704924	S.E. of regression	0.114255
R^2	0.375546	Adjusted R^2	0.340854
$F(3, 54)$	10.82518	P-value(F)	0.000011
Log-likelihood	45.59469	Akaike criterion	−83.18938
Schwarz criterion	−74.94760	Hannan--Quinn	−79.97904
$\hat{\rho}$	−0.083380	Durbin's h	−1.513852

Model 2: OLS, using observations 2001:4–2015:1 ($T = 54$)
Dependent variable: d_l_Inv_SEED_sa

	Coefficient	Std. Error	t-ratio	p-value
const	0.00362818	0.0159944	0.2268	0.8216
d_COST	−2.46812e–006	9.15367e–007	−2.6963	0.0099*
d_NASDAQ	8.90996e–005	9.91441e–005	0.8987	0.3737
d_IR	0.0350396	0.0488400	0.7174	0.4769
d_IR_1	0.0929384	0.0423809	2.1929	0.0336*
d_IR_2	0.00391752	0.0448277	0.0874	0.9308
d_RDGDP	0.747376	2.92366	0.2556	0.7994
d_RDGDP_1	1.89654	2.56678	0.7389	0.4639
d_l_Inv_SEED_sa_1	0.111864	0.117562	0.9515	0.3465
d_l_Inv_SEED_sa_2	0.344820	0.107999	3.1928	0.0026*

Mean dependent var	0.000745	S.D. dependent var	0.125168
Sum squared resid	0.453758	S.E. of regression	0.101551
R^2	0.453532	Adjusted R^2	0.341755
$F(9, 44)$	4.057454	P-value(F)	0.000770
Log-likelihood	52.41508	Akaike criterion	−84.83015
Schwarz criterion	−64.94031	Hannan--Quinn	−77.15942
$\hat{\rho}$	−0.056004	Durbin's h	−0.817113

Model 3: OLS, using observations 2001:1–2015:4 ($T = 60$)
Dependent variable: d_l_Inv_LATER_sa

	Coefficient	Std. Error	t-ratio	p-value
const	−0.0151957	0.0216310	−0.7025	0.4852
d_COST	−2.02616e–006	1.44875e–006	−1.3986	0.1673

Mean dependent var	−0.011086	S.D. dependent var	0.167339
Sum squared resid	1.598241	S.E. of regression	0.166000
R^2	0.032623	Adjusted R^2	0.015945
$F(1, 58)$	1.955973	P-value(F)	0.167269
Log-likelihood	23.62692	Akaike criterion	−43.25385
Schwarz criterion	−39.06516	Hannan--Quinn	−41.61542
$\hat{\rho}$	0.081992	Durbin--Watson	1.631446

Model 4: OLS, using observations 2001:4–2015:1 ($T = 54$)
Dependent variable: d_l_Inv_LATER_sa

	Coefficient	Std. Error	t-ratio	p-value
const	−0.00301208	0.0206914	−0.1456	0.8849
d_COST	−1.60167e–006	1.18660e–006	−1.3498	0.1837
d_NASDAQ	0.000302385	0.000128162	2.3594	0.0226*
d_IR	0.0631656	0.0625818	1.0093	0.3181
d_IR_1	0.0599632	0.0543224	1.1038	0.2754
d_IR_2	0.0235756	0.0561391	0.4199	0.6765
d_RDGDP	4.03098	3.78652	1.0646	0.2926
d_RDGDP_1	−5.63800	3.31715	−1.6996	0.0960

Mean dependent var	0.009949	S.D. dependent var	0.146940
Sum squared resid	0.798252	S.E. of regression	0.131732
R^2	0.302435	Adjusted R^2	0.196283
$F(7, 46)$	2.849092	P-value(F)	0.014842
Log-likelihood	37.16383	Akaike criterion	−58.32765
Schwarz criterion	−42.41578	Hannan--Quinn	−52.19106
$\hat{\rho}$	−0.405937	Durbin--Watson	2.798794

Appendix B

Appendix 2

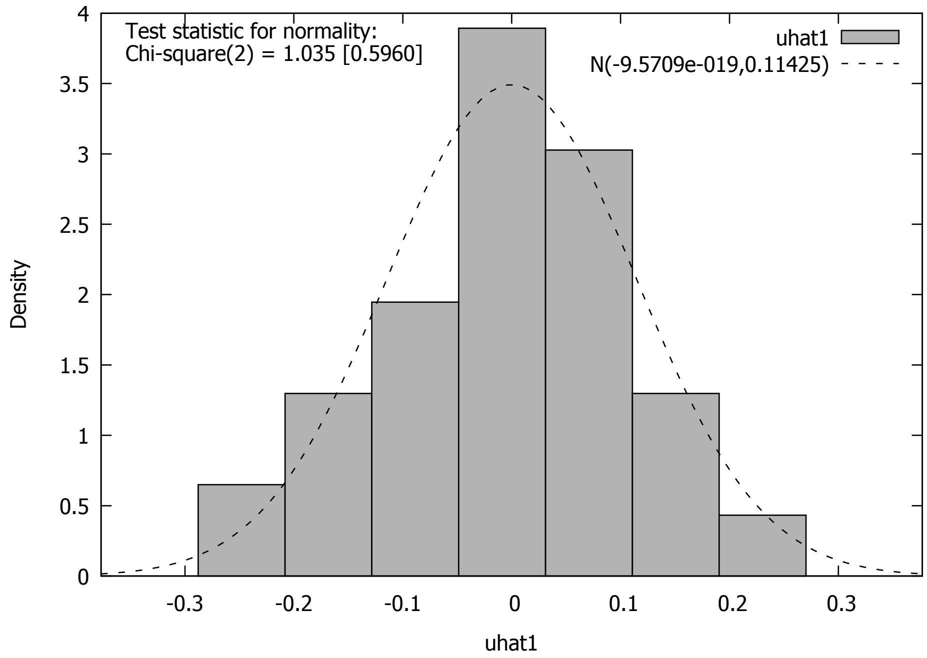

Figure B.1: Residuals plot of Model 1

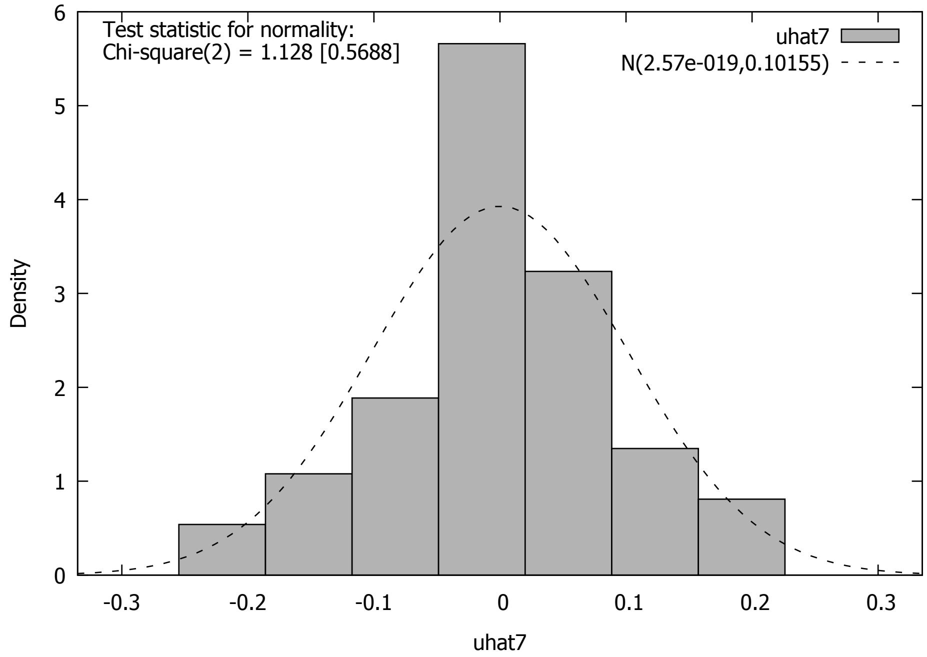

Figure B.2: Residuals plot of Model 2

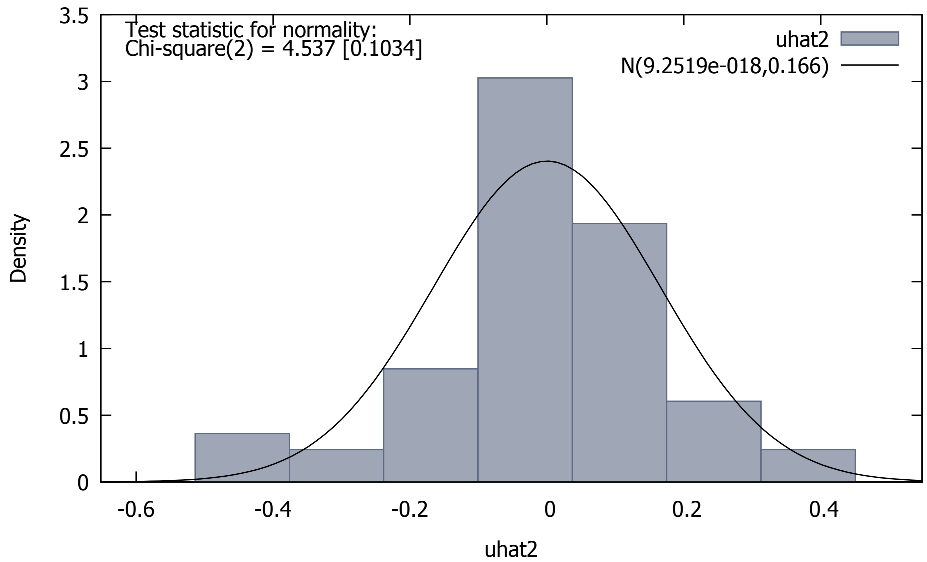

Figure B.3: Residuals plot of Model 3

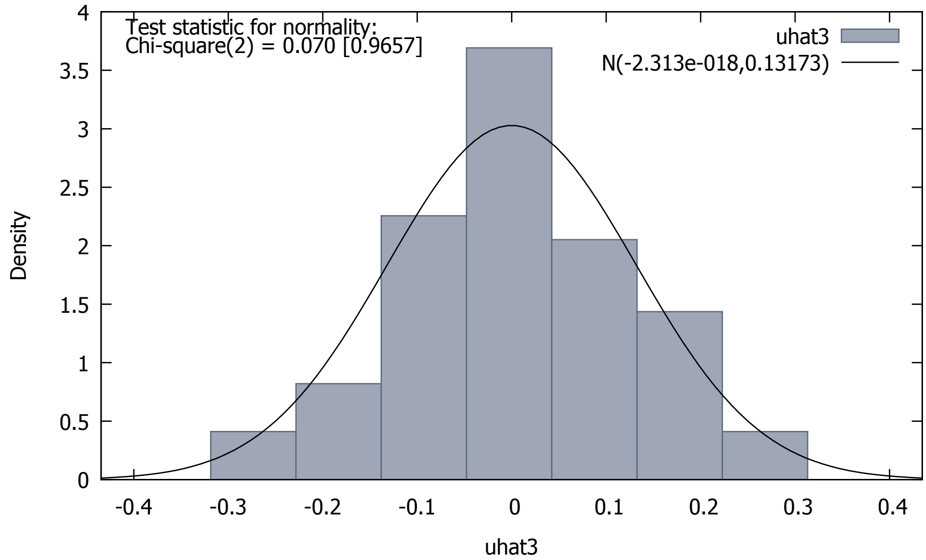

Figure B.4: Residuals plot of Model 4

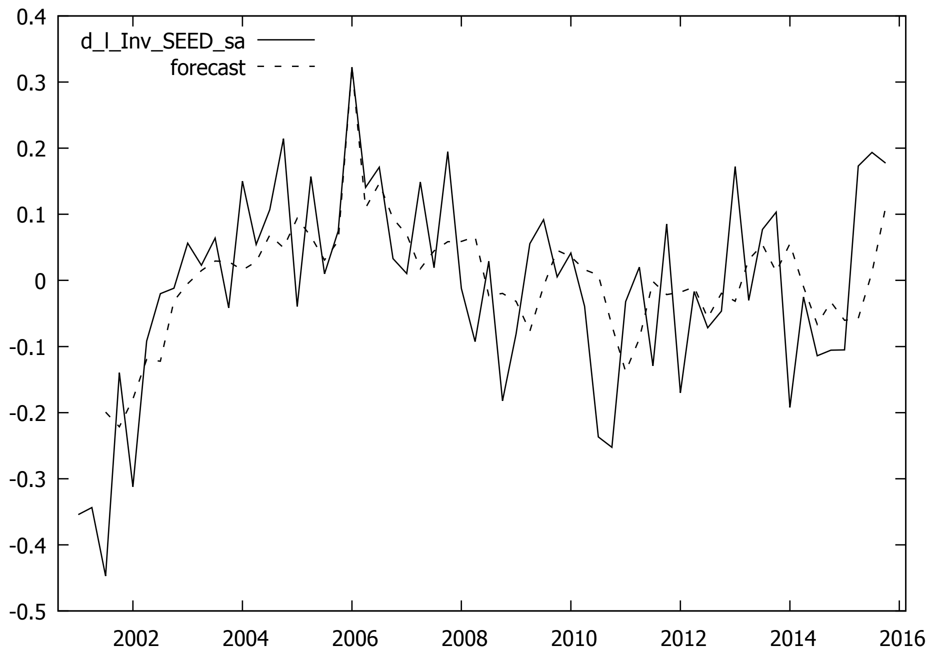

Figure B.5: Forecast of Model 1