Probability of Going Public for Cleantech Startups Based on Fundraising Milestones

Probability of Going Public for Cleantech Startups Based on Fundraising Milestones

Uliana Gorbolskaya

A Thesis in the field of Sustainability and Environmental Managements for the Degree of

Master of Liberal Arts in Extension Studies

Harvard University

November 2015

Abstract

The path to the wide adoption and commercial success of cleantech technologies is hindered by large research and development costs, long testing times, and uncertainty over long-term viability. These risks have made it difficult for many cleantech startups to raise investment capital. Fortunately, the cleantech sector has gained a lot of traction over the past decade, with many companies entering public financial markets, the golden standard for a successful investment exit strategy. This thesis collects and examines empirical data on the private and public investment into the cleantech sector between 2000 and 2015. The data contains information on nearly 4,000 investment rounds received by over 1,000 startups from the top ten cleantech nations and the six oldest cleantech sectors. I fit a statistical model to predict the probability of a company going public based on the different investment sources the company was able to attract. I prove that not all financial sources have equal significance, and only loans, structured debt, and Series A are predictors of a company going public.

Acknowledgements

I wish to express my deepest appreciation to all those who offered their assistance, guidance, and support in this research project. My research advisor, Mark Leighton, PhD, enthusiastically supported the use of statistical methods to analyze investment activity in the cleantech ecosystem.

This research project would not have been successful without the guidance and thoughtful direction of my thesis advisor, Ramon Sanchez, ScD, as well as his courteous offer to provide access to the proprietary Cleantech Group i3connect startup database as the basis for this thesis. Troy Ault, director at the i3 department of the Cleantech Group, also gave crucial guidance into the system and generously provided further data when needed.

I am indebted to the faculty of the Sustainability and Environmental Management program of the Harvard University Extension School, who helped guide my understanding of the relationship between financial, technological and environmental problems. Last, but certainly not least, I thank my friends and family, particularly Tatiana and Oleg Gorbolskiy, for their patience, understanding, and unconditional support.

Table of Contents

List of Tables

List of Figures

Definition of Terms

Cleantech: a broad complex of industry technologies dealing with energy generation, efficiency, storage and infrastructure, waste treatment, transportation, agriculture, materials, manufacturing, and water management (Parad & Cleantech Group, 2014). For this paper I narrowed down the list of subsectors to solar, wind, energy efficiency, energy storage, biofuels, and hydropower.

Growth equity: private equity investment into a company in its expansion phase. Investing in growth funds requires a tolerance for risk and a holding period with a time horizon of five to ten years (Investopedia, 2015a).

Loan guarantee: a promise by one party (the guarantor) to assume the debt obligation of a borrower if that borrower defaults. A guarantee can be limited or unlimited, making the guarantor liable for only a portion or all of the debt (Investopedia, 2015b).

Project finance: the financing of long-term infrastructure, industrial, and public service projects based upon a non-recourse or limited-recourse financial structure where project debt and equity used to finance the project are paid back from the cash flows generated by the project (Investopedia, 2015d).

Risk finance: all forms of financing other than traditional bank loans (Criscuolo & Menon, 2014).

Seed capital: the initial capital used to start a business. In this study, seed capital refers to a small, early-stage equity investment, before a formal Series A round.

Series A: the first formal round of institutional investment. Generally, this is the first time that company ownership is offered to external investors. Series A financing may be provided in the form of preferred stock and may offer anti-dilution provisions in the event that further financing through preferred or common stock occurs in the future (Investopedia, 2015e).

Series B: the second round of financing for a business by private equity investors or venture capitalists. Successive rounds of financing or funding a business are termed Series A, Series B (and so on) financing. The Series B round will generally take place when the company has accomplished certain milestones in developing its business (Investopedia, 2015f).

Structured debt: a service that generally involves highly complex financial transactions offered by many large financial institutions for companies with very unique financing needs. These financing needs usually don't match conventional financial products such as loans (Investopedia, 2015g).

Technological transfer: the process of transferring scientific findings from one organization to another for the purpose of further development and commercialization.

Venture capital: financial capital provided by institutional investors to early-stage, high-potential growth startup companies in exchange for equity. Venture capital specifically targets companies with novel technology or business models in high-tech industries. In this study, venture capital is an umbrella term for seed capital, Series A, Series B, and growth equity.

Chapter I

Introduction

Changes in global population and consumption have had a profound impact on the energy industry in recent years. In the long term, significant technological change will be required to balance the needs of humanity with the planet's natural resource capacity (Pachauri et al., 2014). This shift will require a range of new products and processes across a varied number of industry sectors, typically referred to as "cleantech." However, the path to the wide adoption and commercial success of these new technologies is hindered by large research and development (R&D) costs, long testing times, and uncertainty over long-term viability (Caprotti, 2012; Parad & Cleantech Group, 2014). These risks have made it difficult for many cleantech startups to raise investment capital. Fortunately, the cleantech sector has gained a lot of traction over the past decade, and some subsectors, such as solar and wind, are now reaching market maturity (International Renewable Energy Agency [IRENA], 2015), and therefore present an interesting opportunity for research.

Research Significance and Objectives

This thesis carries out a large sample study of the cleantech sector over the past 15 years, focusing on whether the availability of financing across different investment types was significant in determining the probability of a company going public.

My research will address the following broad objectives: (a) determine whether different investment types have a significant impact on the probability of a cleantech company going public and (b) measure the contribution of each investment type to the probability of a cleantech company going public.

The topics of clean energy innovation, diffusion, and policy making are widely discussed in the literature (Gort & Klepper, 1982; Jaffe, 2012; Newell, Jaffe, & Stavins, 1998; Verspagen, 2004). However, there are few empirical studies that apply statistical learning methods to inferring relationships between different types of finance in different development stages of a cleantech startup. This study comes at an important time, when the first wave of clean technologies is close to maturity. The stakeholders that have participated in this wave of innovation can learn from the previous experience and create better-informed strategies and policies.

This thesis is organized as follows: Chapter I describes the innovation financing cycle and challenges of cleantech investments, summarizing the debate on the role of venture capital (VC) in this subsector. Chapter II describes the methodology and discusses the statistical framework, highlighting the assumptions and possible limitations of the analysis. Chapter III describes the sample and presents some interesting patterns in the data. This chapter also reports estimation results and interprets the statistical output. Chapter IV discusses the findings, highlighting both expected and unexpected results, and provides recommendations and conclusions.

Background

Lubin and Esty (2010) name sustainability one of the major industry megatrends of the current decade. Globalized workforces and supply chains have created environmental pressures that are increasing the competition for natural resources between the rising world powers, primarily China and India (Lubin & Esty, 2010). For years, the international community has seen clean energy innovation as a solution to impending climate change. The Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report (AR5) notes that there is, "high agreement and much evidence that all stabilization levels (of total anthropogenic emissions) assessed can be achieved by deployment of technologies that are either currently available or expected to be commercialized in coming decades." AR5 finds that renewable energy has a big role to play in transforming the energy supply sector, which is the largest contributor to global greenhouse gas (GHG) emissions at 35% of total emissions (Pachauri et al., 2014).

In literature, the concepts of industry evolution were first described by the signature work of Gort and Klepper (1982), who showed that any industry goes through five stages in its lifecycle. It begins with Stage I, when one or more major innovations by the product's first producer are successfully commercialized. In Stage II, the industry experiences a sharp increase in the number of producers and total industry output, together with a fall in output price, particularly towards the end of this phase. Finally, an industry enters a maturity phase (Stage III-V), often through a shake-out-like process, during which the number of producers sharply declines and then stays constant, and both output growth and price declines are much slower (Gort & Klepper, 1982).

Indeed, certain subsectors of cleantech are well on their way to market maturity: solar photovoltaics (PV) are leading the cost decline, with solar PV module costs falling 75% between 2009 and 2015 and the cost of electricity from utility-scale solar PV falling 50% between 2010 and 2015 (IRENA, 2015). In 2014, biomass, hydropower, geothermal, and onshore wind were all competitive with or cheaper than coal, oil, and gas-fired power stations, even without financial support and despite falling oil prices (IRENA, 2015).

The cost reductions have accelerated the diffusion of cleantech technologies in the energy market. Bloomberg New Energy Finance (BNEF, 2015) predicted that carbon-zero energy sources would represent over half of world capacity by 2040. This prediction was based purely on the improving economics of the renewables; there is no government support factored in starting from 2018 (and 2030 for wind), and there only incremental improvements expected on existing technologies such as wind, solar, and energy storage (BNEF, 2015a).

With that, this older generation of renewables is an important step to low-carbon economy; however, it is only part of the equation. To truly push the world towards sustainability, we will need even better and cleaner solutions (Nordan, 2013), which means new waves of innovation need to happen. In the next five years, we are likely to witness rapid development of new cleantech subsectors. Unconventional natural gas production, electric vehicles, advanced internal combustion engines (ICE), and light-emitting diode (LED) lighting are all on the verge of market maturity (Rogers, 2012).

Development Stages and the Valley of Death

However, commercialization remains challenging for cleantech ventures. These challenges are evident during a development phase often referred to as the valley of death: the gap between early-stage, pre-commercial testing and large-scale deployment as seen in Figure 1 (Lester, 2009; Ogden, Podesta, & Deutch, 2008).

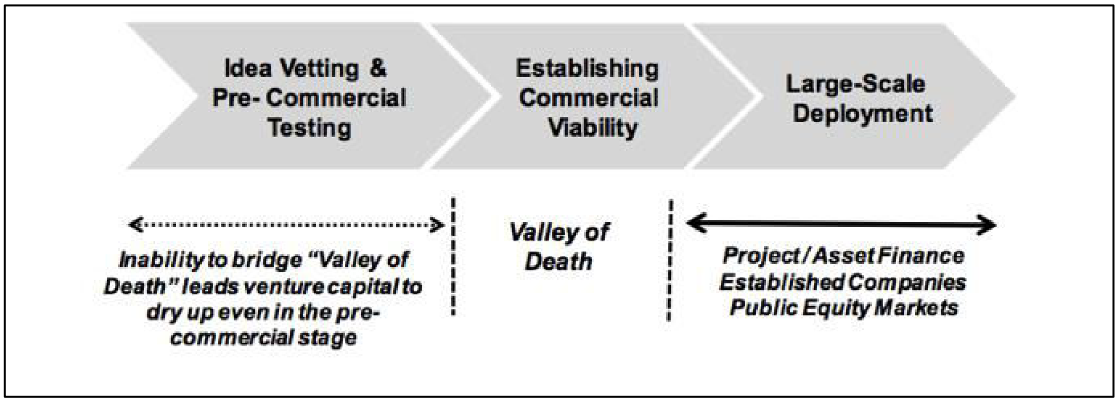


Figure **1**. Funding gaps and the valley of death (Ghosh & Nanda, 2010).

During this phase, financial difficulties are closely related to technological, scalability, and managerial challenges.

Technological and Scalability Valley of Death

Cleantech startups, particularly in the energy production sector, are unique in that they face technology risks at two stages of their commercialization. First, they need to prove that their technology works and then prove that it works at scale (Ghosh & Nanda, 2010). This means that cleantech startups not only remain in the valley of death longer, but they also need more capital investment to demonstrate their technology at scale and provide first proof of commercial pilot.

Managerial Valley of Death

It is typical for founding teams to combine individuals with strong technical and business backgrounds. However, going through all the steps to bridge the valley of death is difficult for managerial teams, as it requires a unique and rare set of skills (Ghosh & Nanda, 2010).

Due to these challenges, the cleantech sector relies heavily on capital availability; however, the traditional order of risk finance type based on startup development stage (Figure 2) does not work as well for cleantech.

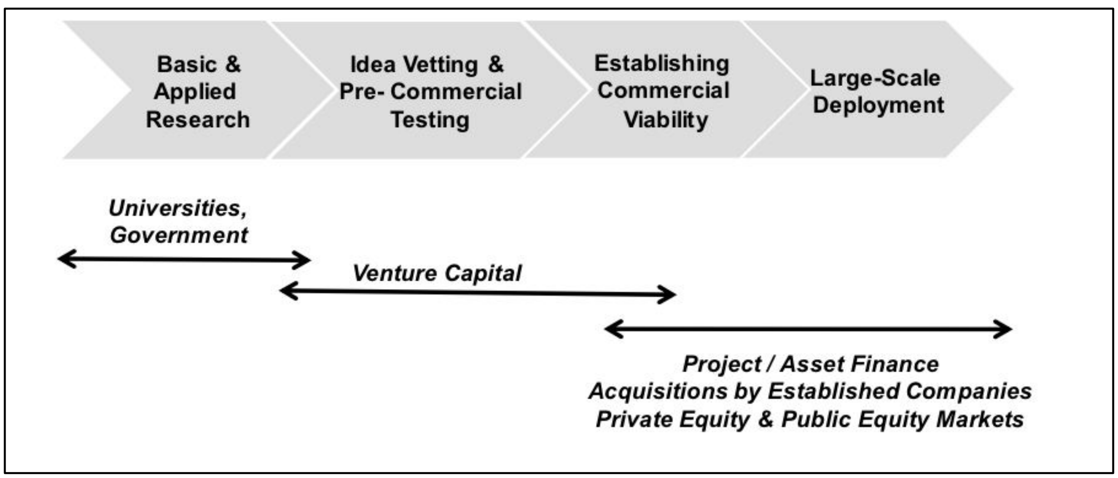


Figure 2. Stage of venture capital investments (Ghosh & Nanda, 2010).

In contrast to the previous successful technological breakthroughs (like information technology), companies in many areas of cleantech can have a higher technological risk profile and simultaneously be more capital intensive, a problem that is exacerbated by uncertain commercial viability (particularly in the short term) and unclear exit strategies (Criscuolo & Menon, 2014). These projects can therefore have a very hard time securing funding.

Ghosh and Nanda (2010) provide a visualization of the risk vs. capital intensity preferences of different sources of risk finance (Figure 3).

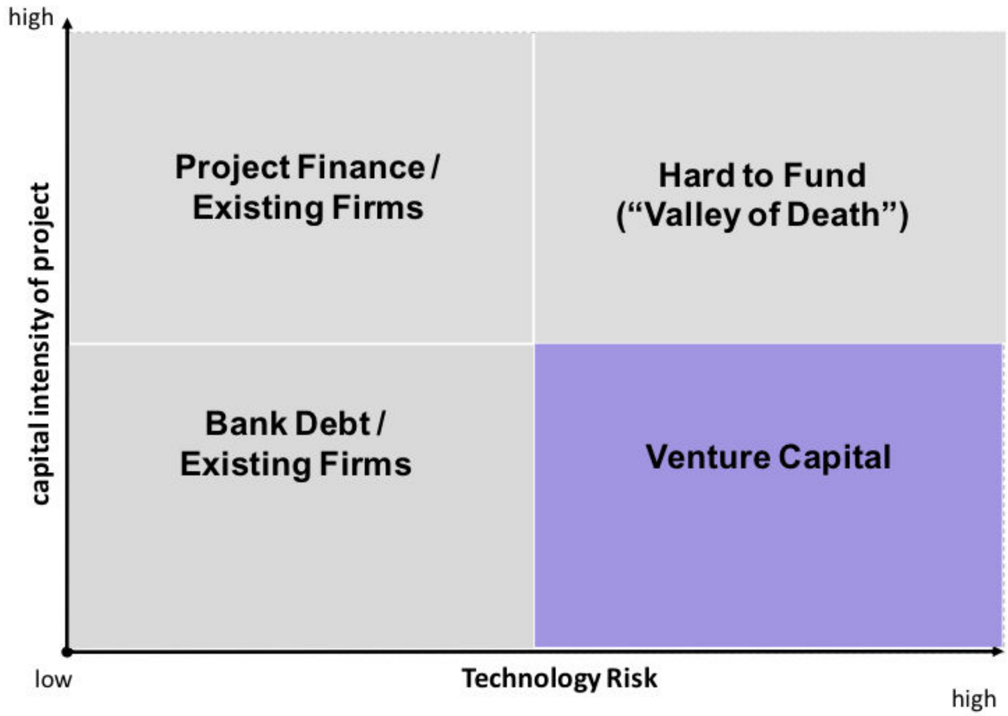


Figure 3. Focus of venture capital investments (Ghosh & Nanda, 2010).

Following this classification, different cleantech technologies (particularly those in later stages of development) can be matched with their most likely source of financing (Figure 4).

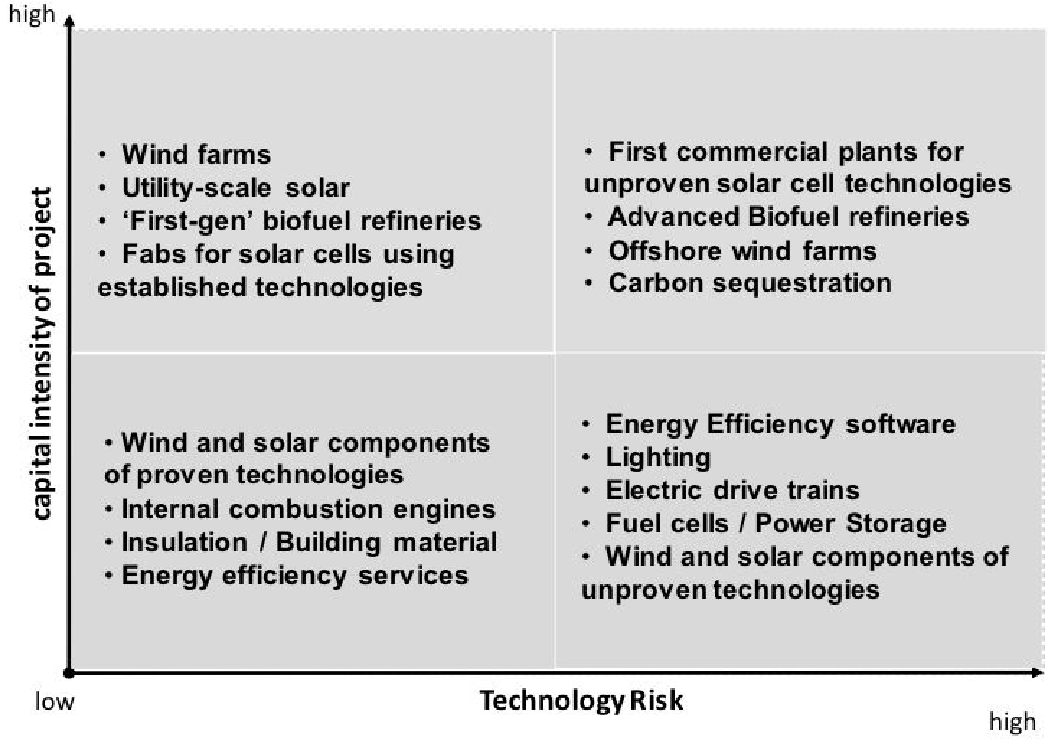
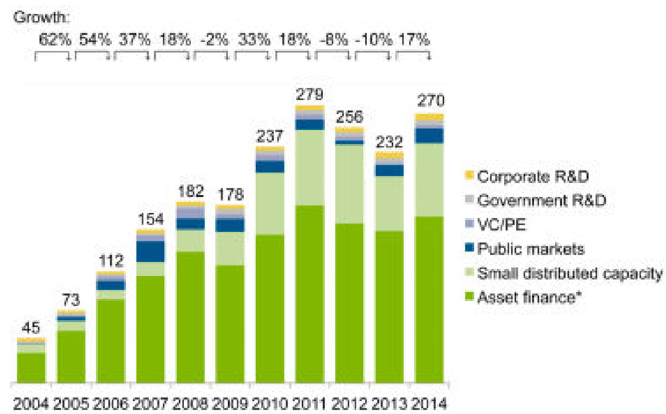


Figure 4. Sub-sectors within clean energy (Ghosh & Nanda, 2010).

As exemplified in the typology outlined in Figure 4, venture capital funds tend to focus on companies that have a high-risk profile but low project capital requirements. Bank debt might be a more appropriate source of funding for projects with low capital needs and low risk profiles, while project finance better suits projects with high capital intensity and lower risk (Ghosh & Nanda, 2010; Kerr & Nanda, 2009).

Deeper Look at Clean Energy Financing

Historically, the lion's share of investment into clean energy technologies comes from asset finance (which includes project finance among other assets) and small distributed capacity (Frankfurt School-UNEP Centre & BNEF, 2015). Only a sliver of world investment comes from corporate R&D, government R&D, and venture financing, although these tend to focus more on early-stage companies and can have a significant impact on the whole picture (Figure 5).



*Asset finance volume adjusts for re-invested equity. Total values include estimates for undisclosed deals
Source: UNEP, Bloomberg New Energy Finance

Figure 5. Global new investment in renewable energy by asset class, 2004-2014, $BN (Frankfurt School-UNEP Centre & BNEF, 2015).

Grants. Governments play an important role with grants and prizes. Howell (2014) showed that public grants, especially at early stages of venture development, approximately double the probability that a firm will receive subsequent venture capital and have large, positive impacts on patenting and the likelihood of achieving revenue. Moreover, it is "the grant money itself that is valuable, not the certification effect, possibly because it funds proof-of-concept work that reduces investor uncertainty about the technology" (Howell, 2014).

The US government supports cleantech innovation through its National Renewable Energy Laboratory (NREL), as well as funding very early-stage investments via special programs, such as Advanced Research Projects Agency-Energy (ARPA-E). China also has a number of government-related green R&D programs, such as the

National High-Tech R&D program (USD 2.9 billion) and the National Basic Research Program, with funding of USD 585 million (Parad & Cleantech Group, 2014). Environmental technology research institutes and laboratories can be found at several Chinese universities (Parad & Cleantech Group, 2014). Philanthropists are also increasingly playing a role in this phase (Criscuolo & Menon, 2014).

Loans and loan guarantee programs. Policy makers can help green-sector entrepreneurs during the scaling-up phase. One example is the US Department of Energy Loan Guarantee Program (LPG), designed to support the development of early stage clean energy (BNEF, 2015b). Under loan guarantee programs such as the LPG, the federal government agrees to cover the debt obligation should a borrower default. In this way, the US government is able to support the innovative clean energy technologies that are typically unable to obtain conventional private financing due to high technology risks (Wang, 2013).

Venture capital. In recent years, venture capital and private equity (VC/PE) has played an increasingly important role, especially in the United States, the United Kingdom and more recently in China (Figure 6). Venture capital funds generally finance multiple projects that have low capital intensity, can show rapid commercial viability (three to five years), and can be sold within the life of a fund (about ten years).

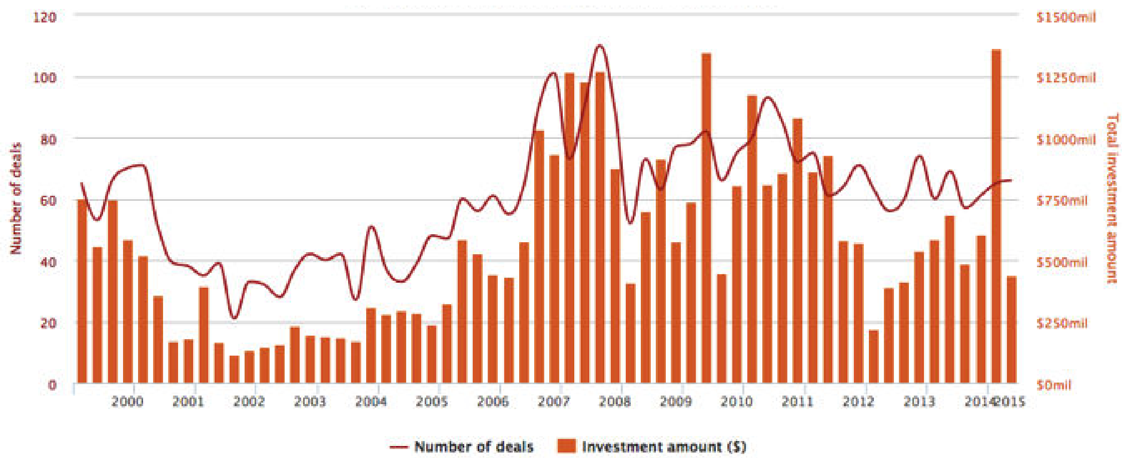


Figure 6. VC investment amount in energy sector, 2000-2015 (Source: PwC, 2015).

Venture capital has been considered a key feature of the successful takeoff of industries, such as IT, software, and biotech. More recently, it is becoming increasingly important for the green takeoff. The risks in the green sector might have particularly important implications for VC financing compared to other sectors such as IT, where VCs have traditionally been very active. These risks include managerial gaps, financing gaps, long horizons, uncertain exits, and regulatory uncertainties (Criscuolo & Menon, 2014). This has been a relatively volatile source of funding for the cleantech startups, with the total number and sums invested changing significantly from year to year (Figure 7).

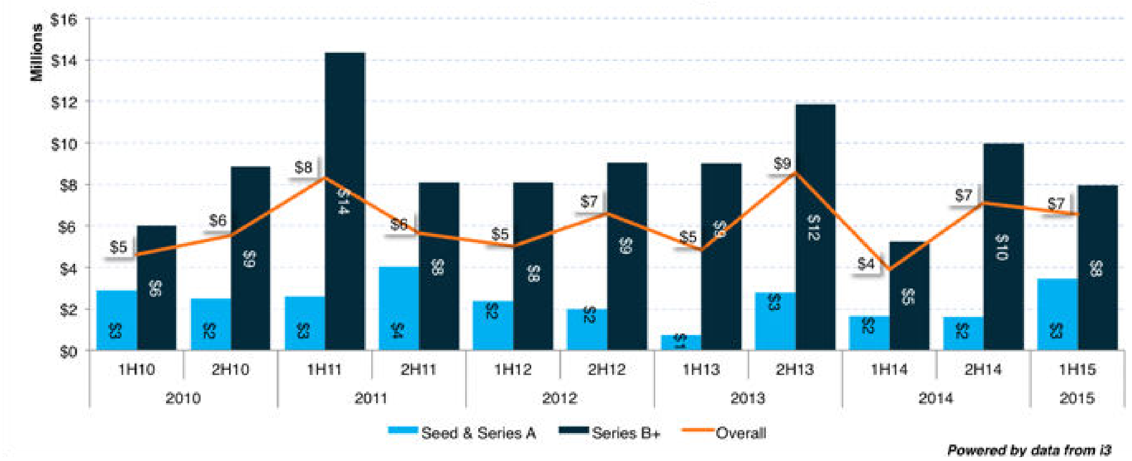


Figure 7. Cleantech venture round average deal size (Cleantech Group, 2015).

To differentiate between various types of venture capital, i3connect (the proprietary Cleantech Group database on cleantech startups) separates them into early stage (seed and Series A), and later stage (Series B and growth equity) groups. Over the last couple of years, there has been a gradual decline in the total number of VC deals in both early and later stages (Figure 8).

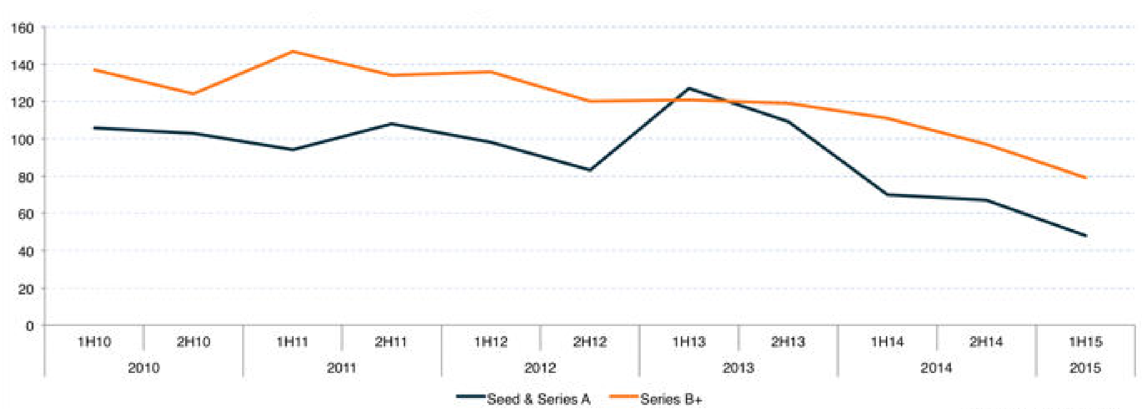


Figure 8. Early vs. late stage venture deal volume (Cleantech Group, 2015).

Structured debt. Over the last several years, green bonds have increasingly attracted attention in global capital markets. Bloomberg New Energy Finance has described these debt securities as “an emerging source for clean energy capital,” and a rapidly growing one at that. In fact, between 1995 and 2013, green bonds were issues for a cumulative $37.8 billion and reached almost the same amount in 2014 alone (Bullard, 2014).

This class of financial assets is rapidly expanding, and The Green Bond Principles were developed as a way to promote integrity within this rapidly growing market (CEG & Croatan Institute, 2014).

Growth Equity. Companies are now delaying their initial public offerings because they’re able to rely on private investors, such as those in the growth equity sector, to help fund their next round of capital. Private funding is now filling their capital needs faster and making companies more profitable by raising their valuations.

In fact, some companies that might have previously gone public in an effort to raise capital are now getting acquired before ever hitting the public market. There are more exits in the form of strategic sales than through initial public offerings (Nordan, 2013).

Project Finance. This type of highly risky and capital-intensive projects are not funded through project finance either, even though this source of financing has been steadily growing since 2004 for projects employing proven clean energy equipment (these projects would be in the top left quadrant of Figure 4). Recent data show that even before the financial crisis, almost no private project finance capital was available for projects whose aim *was* to deploy unproven technologies, and the financial crisis has made capital availability for this type of project even more scarce (Bloomberg New Energy Finance, 2010).

Public Financial Markets. By going public, a public equity market provides an alternative source of finance to banks and venture capital, which is particularly appealing for companies with large current and future investments, high leverage, and high growth (Pagano, Panetta, & Zingales, 1998).

Initial public offering (IPO) is the primary way to access the public equity market. The IPO process is expensive, with the median cost of an IPO in 2013 of $3.3 million (WilmerHale, 2014); therefore, a company would not opt to go through an IPO unless it were confident that it could sell shares at a good price. This generally means a company is viable.

Furthermore, IPOs are underwritten by investment banks that pre-purchase the stock at a discounted price and sell them to institutional investors; both are sophisticated financial intermediaries that would not support an IPO unless they were confident the company was well-managed and was generating a substantial return (WilmerHale, 2014). This makes IPOs an interesting lens to see where the smart money thinks the industry is going.

Research Hypotheses and Specific Aims

The null hypothesis for this research work is that different types of financing do not have a statistically significant impact on the probability of market success. In other words, this thesis tests whether the probability of an IPO is the same for different types of financing (seed, grant, loan, loan guarantee, project finance, Series A, Series B, growth equity, and structured debt).

The main purpose of this study is to infer the underlying relationship between the response variable, going public, and various investment types based on the 2000-2015 data.

The specific aim of the study is to answer the following questions: (a) "Which predictors are associated with the response?" (b) "What is the relationship between the response and each predictor?" (c) "How does each of the individual variables affect the probability of the response variable?" (d) "Can the relationship between market success and each predictor be adequately summarized using linear regression, or is the relationship more complicated?" and (e) "Do the same relationships hold across the top fundraising subset of the sample?"

Given the uncertainty regarding the access to capital markets for cleantech entrepreneurs, this thesis applied statistical learning methods to the development of the cleantech sector in the past 15 years. The goal was to infer relationships between the market exit strategy and the types of financing that these companies have been able to attract. Statistical learning refers to a set of tools for modeling and understanding complex data and encompasses many methods, including regression and classification problems (James, Witten, Hastie, & Tibshirani, 2006).I based my analysis on the i3connect cleantech database, courtesy of the Cleantech Group. I fit a model to predict the probability of a company going public based on the total paid in capital, year founded, sector, geography, and access to various types of finance (seed, grant, loan, loan guarantee, structured debt, Series A, Series B, project finance, and growth equity). As a result, I inferred statistically significant types of finance and the nature of their relationship with market exit. While the final model may be able to predict the probability of going through an IPO for any given startup, it is not my intention to build a model for its predictive power. Instead, the methods applied in this thesis allowed me to achieve most interpretable results. My findings have interesting implications for startup founders, investors, and policy-makers as discussed further in the study.

Chapter II

Methods

Methodologically, this study has two significant components: statistical method selection and data cleaning. These tasks were carried out using R for data analysis and Microsoft Excel for data cleaning.

Statistical Method Selection

The choice of a statistical method for this study was a result of several important considerations. First, the goal of the study (understanding the relationship between the variables) prompted me to select a method with the most interpretable results. To do so, I assumed a linear relationship between the response and predictor variables. While such models are more easily interpretable, they are also likely to produce a higher level of noise (James et al., 2006) since they explain complex problems with a simple model. Such approximations are rare in real life, so I expected the method to have a certain bias (error introduced by simplifying assumptions).

Second, since the response would be a qualitative (binary, or two-class factor) variable, I decided to model this prediction by assigning probabilities to the independent variables. This approach is also known as logistic regression, a common statistical tool that uses a maximum likelihood method to fit the data.

The formula for a logistic regression can be defined as:

$$\log \frac{p(X)}{1 \quad p(x)} \div = {}_0 + {}_1X_1 + \ldots + {}_pX_p$$

where X=(X_1, … , X_p) are independent variables and ${}_0$, ${}_1$, … , βk are coefficients that represent the average change in the response based on a unit change for the independent variable holding the rest of the variables at a fixed value (James et al., 2006). The left side of the equation shows the logit function of the odds.

One of the most attractive aspects of this method is that the assumptions are significantly simpler than for the regular linear model. It does not assume a linear relationship between the covariates and the dependent variables, neither does it assume a normal distribution in the predictor variables or homoscedasticity (Agresti, 2002). However, there are two assumptions that must be met to apply this method. First, as mentioned before, I assumed a linearity of the function in x (James et al., 2006). Second, I assumed that the observations were independent (Hilbe, 2009). Since I included all the i3connect companies that had full profile information, I had no reason to believe that this assumption was false.

It is likely that the model doesn’t match the true outcome exactly (James et al., 2006). If the model is too far from the true outcome, the estimate will be poor. However, if I added too many variables, the model may be over fitted, which means that the model will be too sensitive to noise.

To prepare for accuracy testing, I randomly split the data into two groups. The first sample was used to fit the model, while the second sample was used to test the accuracy of the model.

Building the Statistical Model

I evaluated two different approaches to testing the hypothesis by building and comparing two different models:

Binary predictor variables for investment types. In this model, I used a variable for total paid in capital and then binary variables to identify investment type, country, sector, and year of foundation.

This is the full R formula for this model:

IPO <-glm (exit ~ paidincap + sector + countrycode + year + num_rounds + seed+ grant + loan + loanguarantee + growtheq + structdeb + seriesa + seriesb + projectfin, data=train, family=binomial)

Continuous predictor variables for investment types. Unlike the previous model, I did not use the total paid in capital but rather break up this amount by investment types. As a result, I used continuous variables for the investment amounts in each investment type, country, sector, and year of foundation.

This is the full R formula for this model:

IPO <- glm (exit~ sector + countrycode + year + num_rounds + am_seed + am_grant + am_loan + am_loanguarantee + am_growtheq + am_structdeb + am_seriesa + am_seriesb + am_projectfin, data=train, family=binomial)

Interpretation of Model Results

Once the full model was ready, I optimized each model segment by performing the backward elimination technique, when the predictor variables with the highest p-values were removed one-by-one until the AIC of the model stopped decreasing. After that, I calculated sensitivity and specificity for these models and their respective receiver operating characteristic (ROC) curves. A ROC curve is a plot of the true positive rate against the false positive rate for the different possible cut points of a diagnostic test and demonstrates the tradeoffs between the sensitivity and specificity and the accuracy of the test. The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the test. The closer the curve comes to the 45-degree diagonal of the ROC space, the less accurate the test (UCLA, 2015).

I then selected the best-fit model that had the highest significance levels for its variables, the lowest relative AIC value, and the best fit on the ROC curve.

Finally, I applied the same model to the group of top fundraising companies (top 10%). Since the distribution of the data was skewed to the right, I tested whether the relationship between variables in this group of companies followed a different pattern.

I interpreted the coefficients of the model to analyze the relationship between the probability of going public and the unit increase in the independent variables and discuss the results.

Data Cleaning

This research is based on Cleantech Group's i3connect database that provides access to the proprietary intelligence on 24,000+ companies across all sectors in cleantech (Cleantech Group, 2015)

This database was an immense asset for this study. However, as is usual practice with a database of that size, I needed to carefully examine, clean and validate the data. As a result, the data went through three major iterations: (a) raw data, (b) technically correct data, and (c) consistent data (Jonge & Loo, 2013).

Raw Data

I exported three i3connect csv files with: (a) the profile information on the 23,739 companies stored in the database; (b) a log of 17,424 investment transactions; and (c) a list of 6,744 investor companies.

First, I examined the background information on the startups, including the year of foundation, country, sector, status, number of employees, development stage, and revenue range. It immediately became clear that there were many challenges with the raw data, including inconsistent formatting, missing profile information, and erroneous entries.

The foundation year ranged from 1636 to 2030, since the database included profiles for major universities and companies (the 1636 year referred to the foundation of Harvard University), and two (erroneous) future foundation dates for existing companies. To resolve this challenge, I corrected the "future" dates after checking the company websites and imposed year cutoffs between 2000 and 2014.

To validate the country, I first standardized the way the country information was registered by using the ISO 1366 two-letter abbreviations. Then, I verified that the country code was consistent with the zip code format (i.e., US zip code format was validated to be consistent with a string format consisting of five numbers, while Canadian zip codes had six characters and consisted of both letters and numbers). As a final step, I standardized the state and province information based on the ISO 1366 code and validated that the state addresses were consistent with the country addresses.

To further validate that the dataset included cleantech startups, I used sector and country information. I eliminated any rows that had blanks or "other" as their entry for the sector.

Data visualization of country and sector distribution indicated that a disproportionate number of companies could be identified with a handful of the largest countries and sectors. To provide a basis for a consistent further analysis, I took a subset of the database to include only the six oldest and largest cleantech sectors (solar, wind, biofuels, energy efficiency, energy storage, and hydro) and the ten largest cleantech countries (US, UK, China, Canada, France, Germany, Israel, Netherlands, Sweden, and Italy).

When validating the status variable, I eliminated the rows that had "Bankrupt" or "Out of business" and left the "Acquired," "Private," and "Public" companies (for use at a later stage to validate the dependent variable).

When validating the data, I found out that the variables regarding the number of employees, development stage, and revenue range were rarely updated and therefore inconsistent, so these variables were eliminated from the dataset.

Technically Correct Data

Technically correct data is defined as data of the correct R type that adequately represents the value domain of the variable in the column (Jonge & Loo, 2013). At this stage I had created a "clean" subset of companies that were founded between 2000 and

2014 in six cleantech sectors: solar, wind, biofuels, energy efficiency, energy storage, and hydro. To allow for better comparative analysis, I also limited the country of origin to the ten largest cleantech countries (US, UK, China, Canada, France, Germany, Israel, Netherlands, Sweden, and Italy).

Next, I sourced the information on investments from the investments database. I matched the company names with the investment log data, creating binary variables for each investment type (seed, grant, loan, loan guarantee, Series A, Series B, structured debt, growth equity and project finance). A company would get a "1" if it received at least one round of investment of the respective type and "0" if it did not. Further, I attributed several numerical variables to reflect the number of rounds of every type, as well as the total and average amounts paid (by investment type).

I aggregated this information at a company level in three variables: total paid in capital, total number of rounds, and average paid in capital per round. The dependent variable reflects whether a company had a successful exit through an IPO. To create the dependent variable regarding exit, I used the variable "Ticker" (abbreviations used for a respective company on the stock exchange) to create a binary variable "exit" that would take the value of "1" if the company had gone through an IPO and "0" if not.

To make sure that this captured all the companies in the database that went through an exit, I used the "status" variable, discussed earlier. Recall that I only had three factor levels remaining in the status: private, public, and acquired. I ran a number of logical tests on these variables to make sure that for all observations, the exit variable of the value "0" was equal to factor level "Private" or "Acquired" while the exit value of "1" was equal to "Public." Thirty-two observations didn't meet the logical tests since some companies had been both private and gone through an IPO, and I determined the correct exit value through an online search.

Consistent Data

The resulting dataset had 1,169 observations and 43 variables. To better understand and characterize the data at hand, I applied two approaches:

1) I examined the shape and the structure of the data for important features by applying Exploratory Data Analysis (EDA). See the Chapter III for more details.
2) I examined the data for unusual observations (such as extreme values and outliers) through data visualization tools such as scatter plots and box plots (National Institute of Standards and Technology[NIST], 2005).

Extreme Values and Outliers

While an outlier is broadly defined as an extreme value that lies disproportionately far from the other values in a random sample, the pertinent definition of an outlier for a given study has to be aligned with the goals of the study and the distribution of the data in question (NIST, 2005). A multiple logistic regression methodology does not assume a normal distribution of data (as can be seen in more detail further), which gives more flexibility in treating the data.

Given the volatile and unpredictable nature of entrepreneurship and the highly uneven distribution of financial capital, it is not surprising that the data has a lot of extreme values. In all cases but one, I kept these observations in the database. I made the only adjustment in the case of Mingyang Wind Power (Chinese wind giant) that had received over $8 billion from two transactions (a $5 billion structured debt contract and $3 billion in project finance). While this data is accurate, it is more than five times higher than the second largest observations in each respective investment type and hundreds of times above the mean. In this particular case, I felt justified to remove the observation from the dataset.

Chapter III

Statistical Results

The dataset contained 1,169 observations and 43 variables (Table 1).

Table 1. Description of variables for the model.

Variable Name	**Description**	**Variable Type**	**Valid Values**
Exit	IPO	Factor w/2 levels	"0" or "1"
Year	Year	Integer	2000-2014
sector	Sector	Factor w/6 levels	"biofuels," "energy storage," "energy efficiency," "hydro," "solar," and "wind"
countrycode	Country	Factor w/10 levels	"CA," "CN," "DE," "SE," "IT," "FR," "CA," "GB," "IL," "NL," and "US"
paidincap	Total Paid in Capital	Numerical	"0" or positive number
averagecap	Average Paid in Capital	Numerical	"0" or positive number
num_rounds	Total Number of Rounds	Integer	"0" or positive number
Seed	Seed	Factor w/2 levels	"0" or "1"
Grant	Grant	Factor w/2 levels	"0" or "1"
Loan	Loan	Factor w/2 levels	"0" or "1"
loan_guarantee	Loan Guarantee	Factor w/2 levels	"0" or "1"
growtheq	Growth Equity	Factor w/2 levels	"0" or "1"
Strdebt	Structured Debt	Factor w/2 levels	"0" or "1"

(Continued…) Variable Name	Description	Variable Type	Valid Values
Seriesa	Series A	Factor w/2 levels	"0" or "1"
Seriesb	Series B	Factor w/2 levels	"0" or "1"
projectfin	Project Finance	Factor w/2 levels	"0" or "1"
num_seed	Number of Seed Rounds	integer	"0" or positive number
am_seed	Total Amount Received in Seed Rounds	numerical	"0" or positive number
ave_seed	Average Amount per Seed Rounds	numerical	"0" or positive number
num_grant	Number of Grant Rounds	integer	"0" or positive number
am_grant	Total Amount Received in Grant Rounds	numerical	"0" or positive number
ave_grant	Average Amount per Grant Rounds	numerical	"0" or positive number
num_loan	Number of Loan Rounds	integer	"0" or positive number
am_loan	Total Amount Received in Loan Rounds	numerical	"0" or positive number
ave_loan	Average Amount per Loan Rounds	numerical	"0" or positive number
num_loanguarantee	Number of Loan Guarantee Rounds	integer	"0" or positive number
am_loanguarantee	Total Amount Received in Loan Guarantee Rounds	numerical	"0" or positive number
ave_loanguarantee	Average Amount per Loan Guarantee Rounds	numerical	"0" or positive number
num_growtheq	Number of Growth Equity Rounds	integer	"0" or positive number
am_growtheq	Total Amount Received in Growth Equity Rounds	numerical	"0" or positive number

(Continued…) Variable Name	Description	Variable Type	Valid Values
ave_growtheq	Average Amount per Growth Equity Rounds	numerical	"0" or positive number
num_structdeb	Number of Structured Debt Rounds	integer	"0" or positive number
am_structdeb	Total Amount Received in Structured Debt	numerical	"0" or positive number
ave_structdeb	Average Amount per Structured Debt Rounds	numerical	"0" or positive number
num_seriesa	Number of Series A Rounds	integer	"0" or positive number
am_seriesa	Total Amount Received in Series A Rounds	numerical	"0" or positive number
ave_seriesa	Average Amount per Series A Rounds	numerical	"0" or positive number
num_seriesb	Numer of Series B Rounds	Integer	"0" or positive number
am_seriesb	Total Amount Received in Series B Rounds	numerical	"0" or positive number
ave_seriesb	Average Amount per Series B Rounds	numerical	"0" or positive number
num_projectfin	Number of Project Finance Rounds	Integer	"0" or positive number
am_projectfin	Total Amount Received in Project Finance Rounds	numerical	"0" or positive number
ave_projectfin	Average Amount per Project Finance Rounds	numerical	"0" or positive number

General Overview

The 1,169 startups cumulatively raised over $57 billion in almost 4,000 rounds, and 66 companies went through an exit. If asked to define the most typical startup based on the combination of the most common features, this company would be an energy efficiency startup, funded in 2007 in the USA. By 2015 it would have raised approximately $50 million dollars in three financing rounds through a combination of loan, venture capital, and growth equity. To understand the underlying structure and shape of the data, it would be informative to look at the data at a more granular level.

Year Founded, Country and Sector Overview

The year variable had a normal distribution (Figure 9). The data implied that there was a growing trend in the number of companies until 2007, and then the trend seemingly reversed.

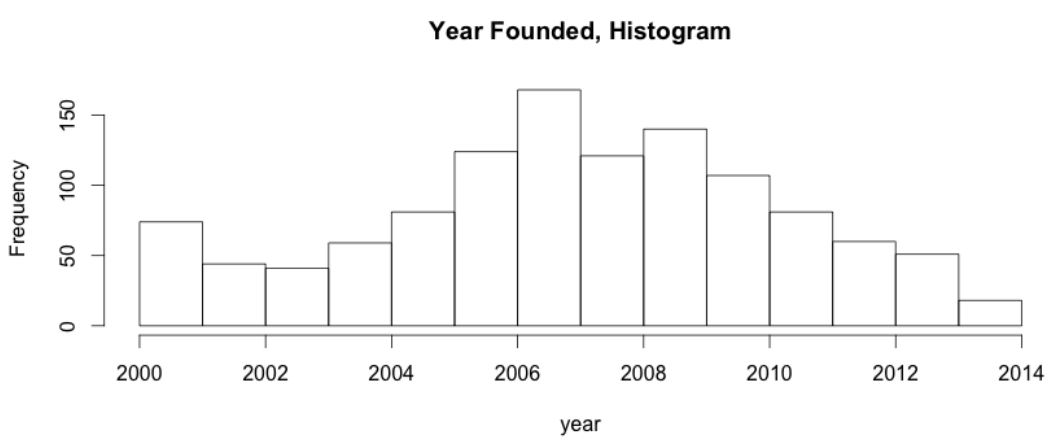


Figure 9. Year founded histogram for cleantech companies.

The composition of startups by country and sector was relatively similar in each of the years, with the majority of startups located in USA and working in the energy efficiency filed (Figure 10).

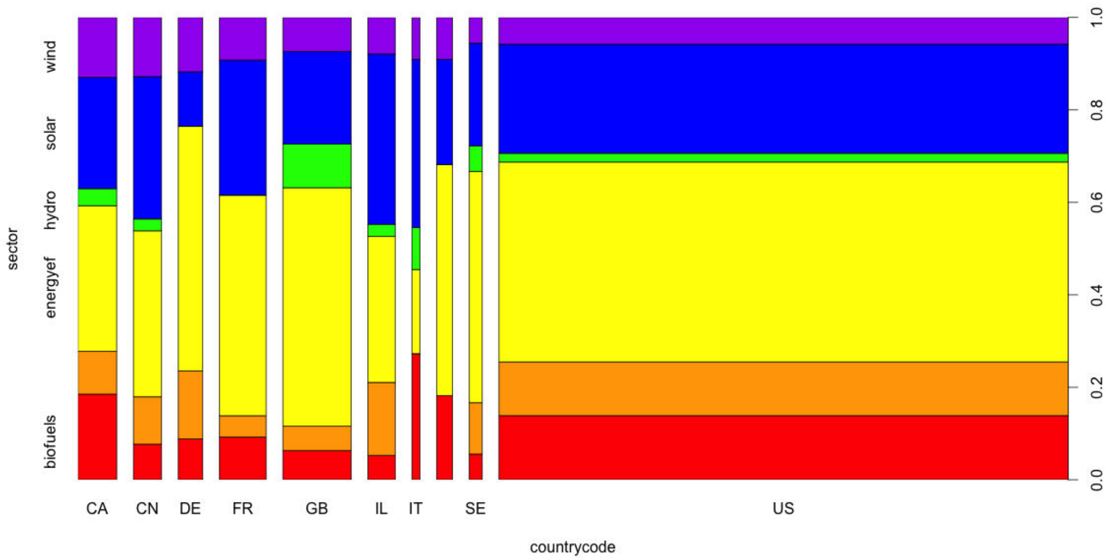


Figure 10. Distribution of companies by sector and country, bar plot.

Sector

Energy efficiency startups represented the largest subsector and 43% of all startups, with solar coming in second (24%) (Table 2). Biofuels, energy efficiency, and energy storage were all around 10%, and hydro represented only 3%. In terms of financing, solar was the leading sector with 46% of total paid in capital, while energy efficiency came in second with 20% and biofuels a close third (18%). In terms of successful exits, through 2014, only 60 companies had gone through an IPO (0.5%). A third of these companies were in the solar subsector, 23% in biofuels, and 20% in energy efficiency.

Table 2. Percentage breakdown of companies, paid in capital, IPOs, and investment rounds, by sector.

	% of companies	**% of paid in capital**	**% of IPOs**	**% of rounds**
Biofuels	13%	18%	23%	15%
Energy storage	10%	8%	9%	13%
Energy efficiency	43%	20%	20%	38%
Hydro	3%	0%	5%	3%
Solar	24%	46%	32%	25%
Wind	7%	8%	12%	5%

The distribution of paid-in capital for any sector was heavily skewed to the right, creating a lot of extreme values (Figure 11). The scale of the Figure 11 was adjusted to provide the most readability of the data; there are more extreme observations beyond this scale.

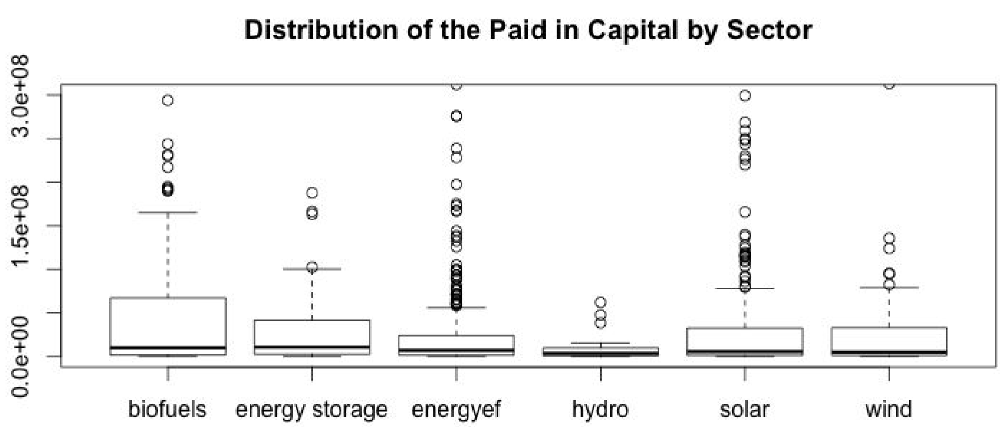


Figure 11. Distribution of the paid in capital by sector, boxplot.

The mean amount of paid-in capital per sector ranged from $92 million in solar to $9 million in hydro (Table 3).

Table 3. Summary statistics of total paid in capital by sector: minimum, mean, standard deviation, maximum.

Sector	Minimum	Mean	Standard Deviation	Maximum
Biofuels	$15,000	$68,579,492	$148,456,591	$1,200,000,000
Energy Storage	$25,000	$36,826,715	$87,542,082	$760,470,048
Energy Efficiency	$0	$22,391,585	$46,566,490	$559,628,426
Hydro	$0	$9,164,585	$14,742,339	$62,221,091
Solar	$0	$92,393,532	$352,018,108	$4,490,000,000
Wind	$50,000	$58,922,059	$19,900,666	$1,170,365,000

In terms of finance composition of the sectors, it is informative to see that a significant percentage of solar and wind capital came through loans and project finance (Figure 12). The most significant source of financing for the other four sectors was growth equity.

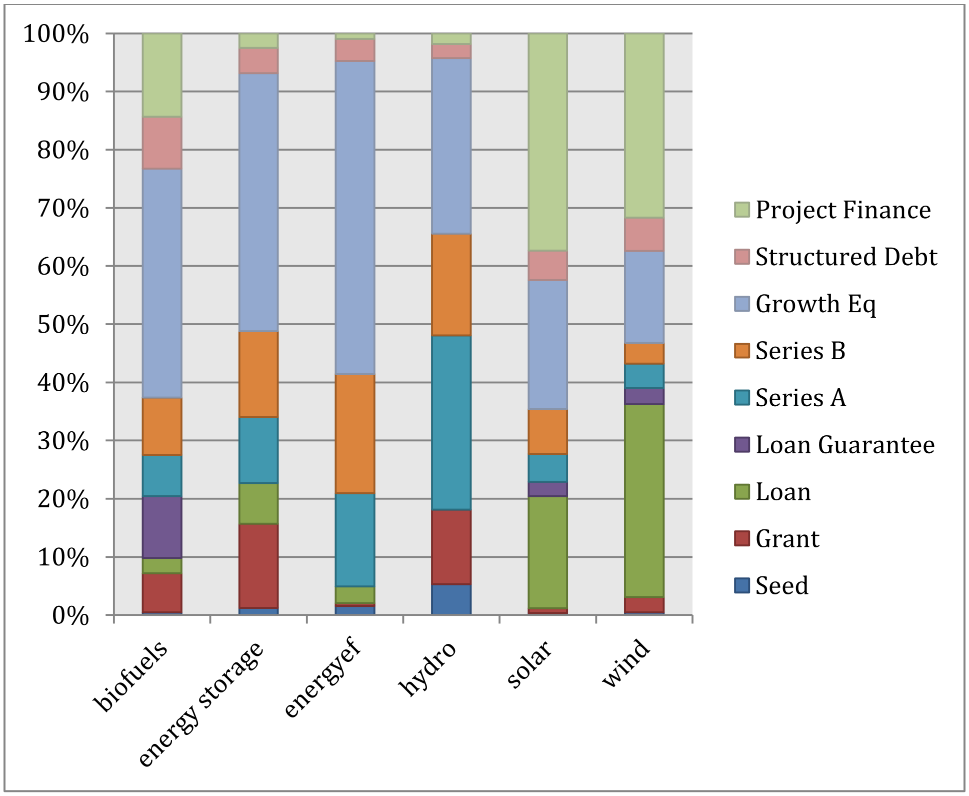


Figure 12. Investment sources by sector, stacked columns.

Country

The majority of the companies were from the US (US - 837 observations), UK (GB - 100), France (FR - 66), Canada (CA - 55), China (CN - 42), Israel (IL - 41), Germany (DE - 39), Netherlands (NL - 22), Sweden (SE - 18), and Italy (IT - 11). The US accounted for 68% of all startups, 78% of all paid-in capital, and 53% of all exits (Table 4).

Table 4. Percentage breakdown of companies, paid-in capital, IPOs, and rounds, by country.

Country	% of companies	% of paid in capital	% of exits	% of rounds
Canada	5%	3%	11%	5%
China	3%	7%	20%	2%
Germany	3%	4%	2%	3%
France	6%	2%	5%	3%
UK	8%	2%	6%	6%
Israel	3%	1%	2%	2%
Italy	1%	0%	0%	1%
Netherlands	2%	1%	0%	2%
Sweden	2%	1%	3%	1%
USA	68%	78%	53%	75%

The distribution of paid-in capital for any country was also heavily skewed to the right, creating a lot of extreme values as reflected in Figure 13 (the y-scale of Figure 12 was adjusted to provide the best readability; there are more extreme values that remained beyond the scale of this figure).

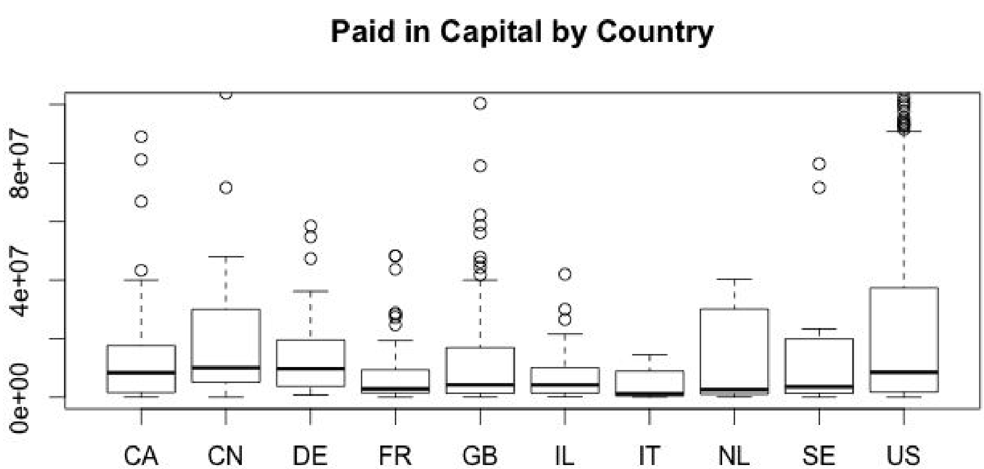


Figure 13. Distribution of the paid-in capital by country, boxplot.

The mean amount of paid-in capital per sector ranged from $106 million in China to $13 million in the UK (Table 5).

Table 5. Summary statistics of total paid-in capital by country: minimum, mean, standard deviation, maximum.

Country	Minimum	Mean	Standard Deviation	Maximum
Canada	$30,700	$32,416,087	$94,906,078	$550,650,000
China	$0	$106,510,369	$347,149,909	$1,950,700,000
Germany	$760,000	$63,747,666	$215,242,158	$1,170,365,000
France	$0	$21,224,152	$96,341,330	$772,028,770
UK	$40,000	$13,458,782	$21,246,371	$112,240,396
Israel	$75,000	$13,811,109	$40,182,465	$248,542,100
Italy	$15,000	$23,443,226	$46,415,666	$117,900,000
Netherlands	$50,000	$20,525,187	$34,899,794	$124,109,000
Sweden	$16,636	$30,059,286	$74,516,842	$313,215,256
USA	$0	$58,182,204	$214,166,359	$4,490,000,000

The countries each have a different risk capital profile (Figure 14). China and Germany led in the share of financing they got from loans, 70% and 55%, respectively. Venture capital (Series A and B) played the leading role in the UK (52%) and Italy (50%), while project finance was a major source in France and Israel, both around 45%. The US and Canada had the most diverse portfolio of investment types.

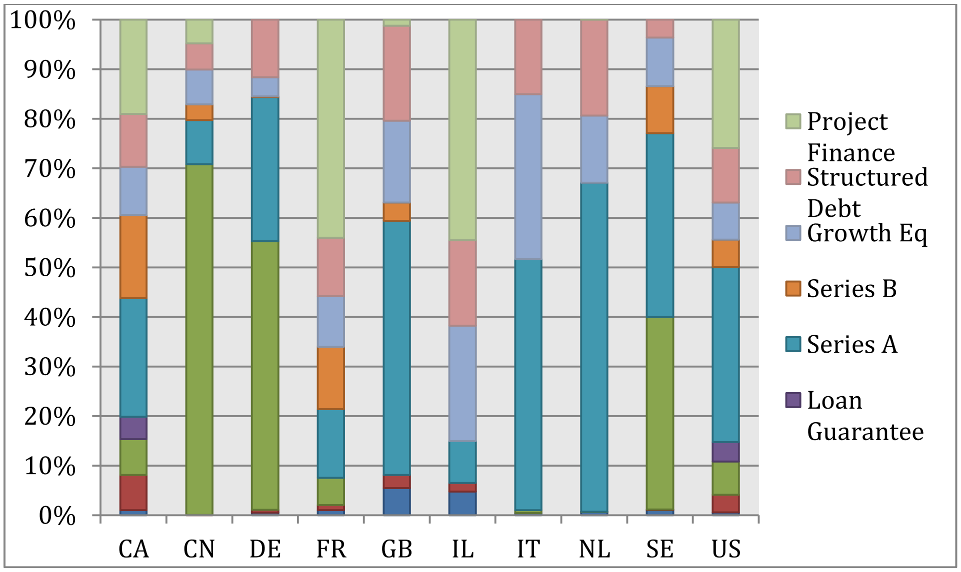


Figure 14. Investment sources by sector, stacked columns.

Investment Rounds

The distribution of data for both the amounts paid and the number of rounds was heavily skewed to the right (Figure 15 and Table 6).

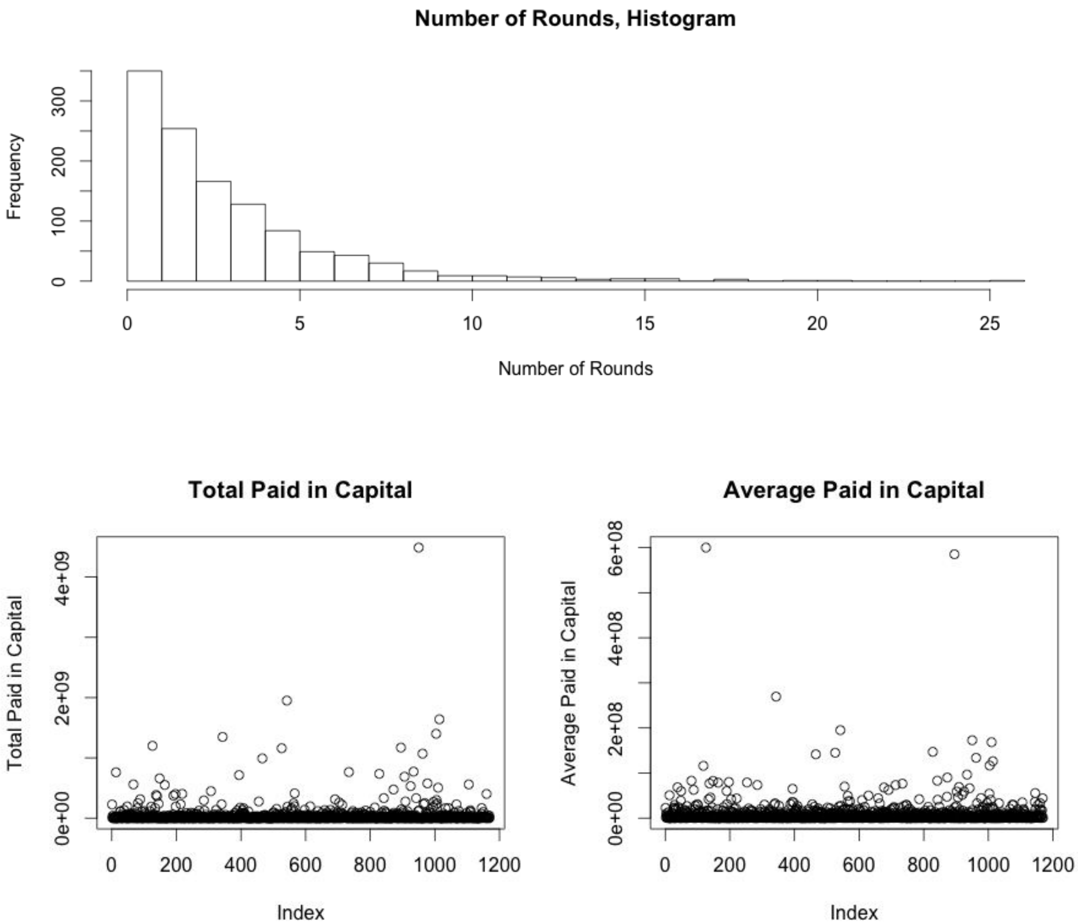


Figure 15. Distribution of number of rounds, total paid-in capital, and average paid-in capital per company.

Table 6. Summary statistics on number of rounds, total paid-in capital and average paid-in capital per company.

	Min.	Median	Mean	Max.
Total Paid In Capital	$0	$6,990,000	48,800,000	$4,490,000,000
Number of Rounds	0	2	3.344	26
Average Paid In Capital	$0	$3,020,000	$12,500,000	$600,000,000

Nine binary factor variables indicated the investment types: seed, grant, loan, loan guarantee, structured debt, Series A, Series B, and growth equity. Almost half of all investment came in the form of growth equity (33%) and project finance (23%). Series A and B together accounted for almost 20%, and loans came up to 13% of all financing.

The final 27 variables expanded on this information: for each investment type, I knew (a) how many rounds of that investment type a company was able to attract (Table 7), (b) total amount paid (Figure 16), and (c) average amount per round (Table 8).

Table 7. Count of total investment rounds by type.

Type of round	**Number of companies**
Seed	424
Grant	216
Loan	48
Loan guarantee	14
Growth equity	499
Structured debt	197
Series A	748
Series B	449
Project finance	49

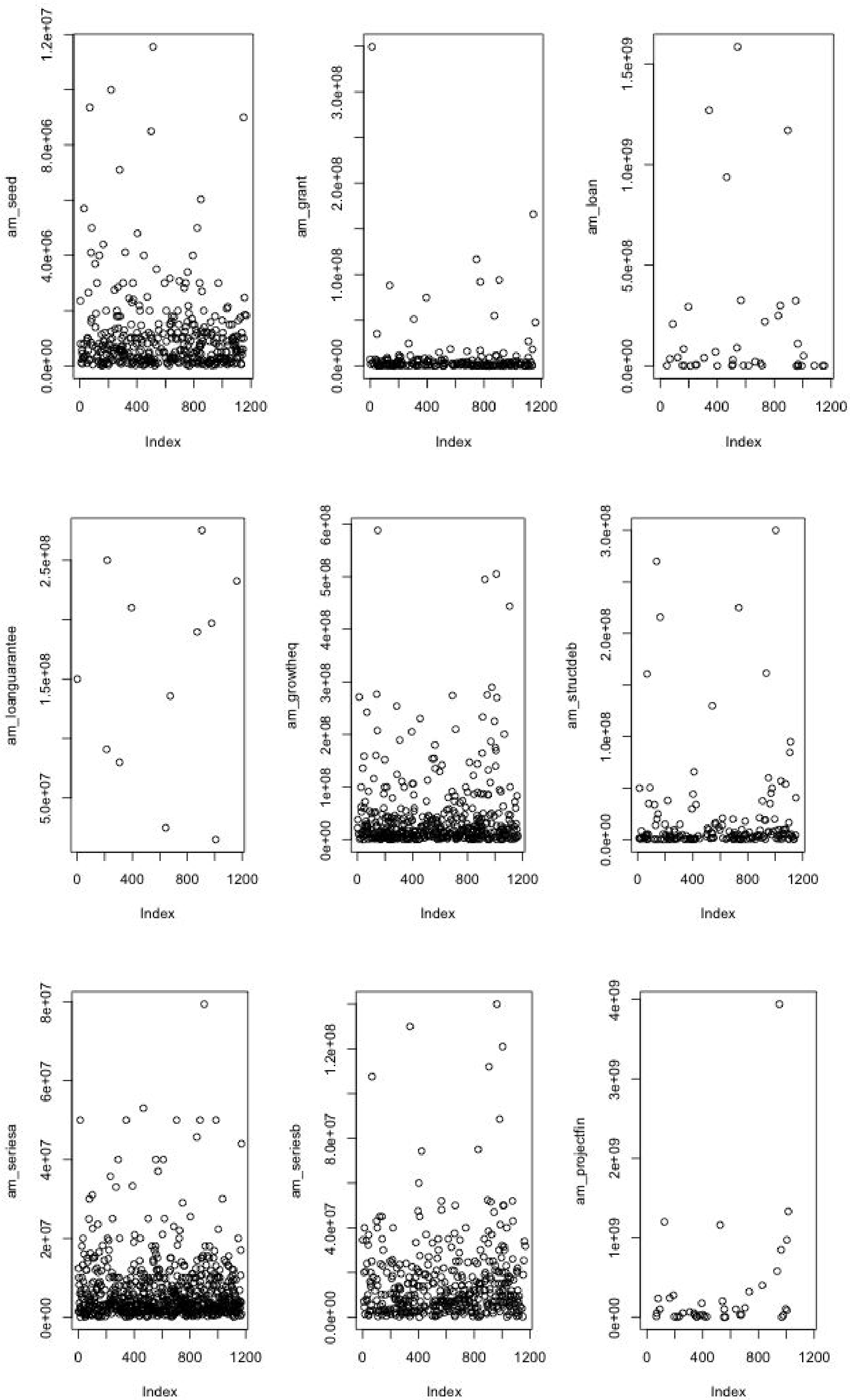


Figure 16. Distribution of total investment amount per type.

Table 8. Summary statistics on various types of financing: minimum, median, mean, maximum.

Investment type	**Minimum**	**Median**	**Mean**	**Max.**
Seed	$0	$0	$337,254	$11,555,050
Grant	$0	$0	$1,517,779	$349,270,048
Loan	$0	$0	$6,425,000	$1,586,000,000
Loan Guarantee	$0	$0	$1,583,199	$275,000,000
Series A	$0	$1,250,000	$3,913,961	$79,400,000
Series B	$0	$0	$5,316,906	$140,000,000
Structured Debt	$0	$0	$2,674,505	$300,000,000
Growth Equity	$0	$0	$15,989,561	$587,872,515
Project finance	$0	$0	$11,040,000	$3,940,000,000

Modeling Exit Probability

I used R to build and optimize the two models, as discussed in the methodology section (see the Appendix for the regression summary). The model with the binary variables showed the best fit based on AIC number (Table 9), though no significant difference was determined based on the sensitivity/specificity test or the ROC curves for either model (Figures 17-20). Based on the AIC fit, I selected this model for further analysis.

Binary Variable Model

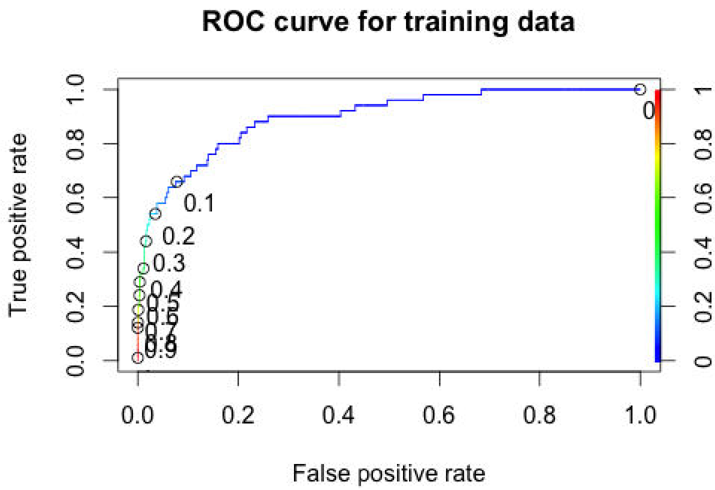


Figure 17. ROC curve for training data, binary variable model.

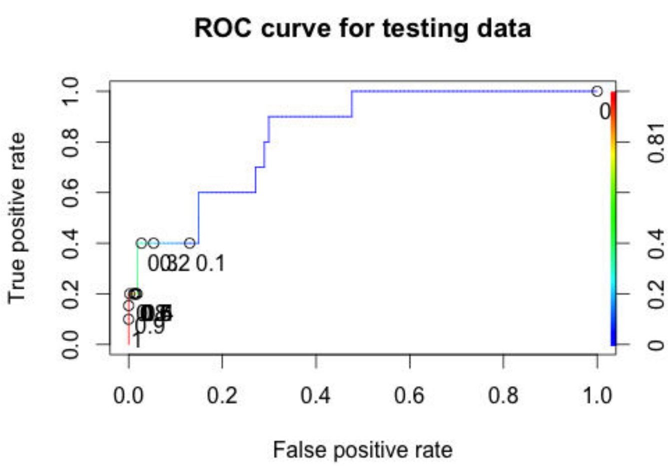


Figure 18. ROC curve for testing data, binary variable model.

Continuous Variable Model

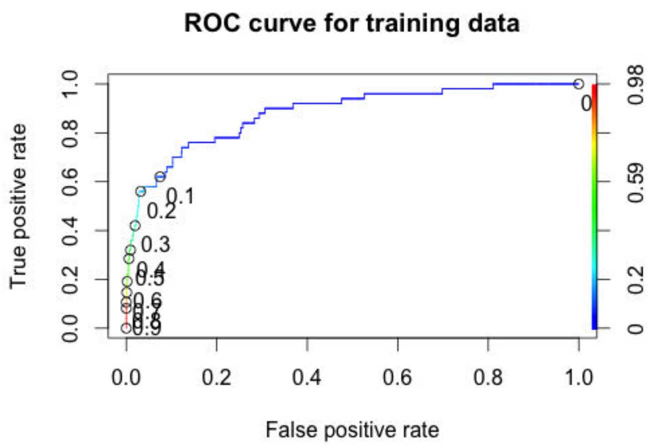


Figure 19. ROC curve for training data, continuous variable model.

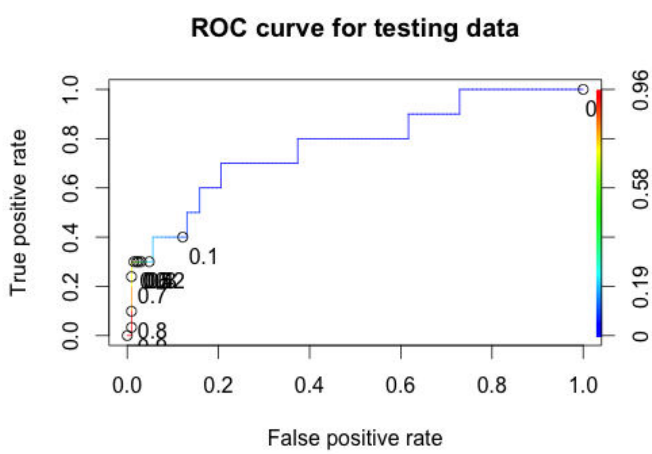


Figure 20. ROC curve for testing data, continuous variable model.

Table 9. Comparative analysis of the two models.

Model:	**AIC:**
Binary investment variable model	297.39
Continuous investment variable model	307.97

Interpreting Model Output for Binary Model

The backward elimination technique on the full binary logistic model yielded the following best fit model:

glm(formula) = exit ~ paidincap + sector + countrycode + year + num_rounds + grant + loan + seriesa + projectfin, family = binomial, data = train)

There were several independent factor variables in the model, and since R automatically calculates the first alphabetized level as the baseline model, the baseline output was calculated for the biofuels sector in Canada. Paid-in capital had a significant relationship with the outcome, as did year, loan, structured debt and Series A financing. Among the multilevel factor variables (sector and country), there was a significant relationship with the energy efficiency sector as well as the country levels for China, UK, and US (Table 10).

Table 10. Interpreting coefficients: significance in the model and relationship with the response variables.

Variable	**p-value**	**Relationship**
paidincap	0.003975 **	Positive
energy efficiency	0.023763 *	Negative
countryCN	0.009212 **	Positive
countryUSA	0.006536 **	Positive
year	0.000275 ***	Negative
loan	0.002137 **	Positive
strdebt	0.029593*	Positive
series a	0.006232**	Positive

Significance codes: 0.001 ‘**’ 0.01 ‘*’

Further, I applied this model against the subset of the top 10th percentile of the fundraising companies. The model output shows no significant variables to predict the IPO exit among this group.

It is interesting that the model containing the absolute values of the investments had a lower predictive power than the model with binary investment types, which suggests that it is the type of investment specifically (not just the amount raised in that round) that is significant. This may indicate that there are unobserved variables behind the investment types that impact the probability of a company accessing the public markets.

Chapter IV

Discussion, Recommendations, and Conclusions

The study suggests several important findings:

- different sources of financing hold different statistical significance for the probability of going public;
- loans, structured debt and Series A financing is significant at the 95% confidence level;
- these findings are significant for biofuels and energy efficiency sectors;
- these findings are significant for USA, China, and Canada; and
- the top 10 percent of the companies did not have any significant correlations between the probability of going public and different sources of finance.

Discussion

The results of the regression analysis show that different sources of financing vary in statistical significance, which disproves the null hypothesis. The positive coefficient for the total paid in capital makes intuitive sense: the more funding a startup is able to attract, the higher probability of an IPO. Similarly, the higher the foundation year (i.e., the younger the company), the lower the probability of an IPO.

Significance of Different sources of financing

In terms of the investment types, the results are not as straightforward. The fact that a company received venture capital (Series A) is indicative of the certainty of the institutional investors in that particular company. However, no later stage investment showed any significant relationship. This can partly be explained through a certain effect of multicollinearity (there is a medium level of correlation (0.5) between Series A and Series B investment). However, the correlation between Series A and growth equity is practically non-existent (0.2). Intuitively, the later stages of venture capital investment should indicate an even more statistically significant relationship with the response variable. This didn’t happen. One possible explanation is that venture capital has been showing an increasing preference for mergers and aquisitions M&A exits over IPO (WilmerHale, 2014).

The final two significant variables—loans and structured debt—are types of financing that a company can access once it is able to repay these debts from its cash flows. This means that the technology is showing traction and has eliminated most of the risks (both technological and commercialization risks) connected with the innovation. In that sense, these sources of financing indicate a later-development stage of a startup and are stronger determinants of an IPO than later-stage venture financing.

Environmental, social, and economic implications. These finding show that the choice of financing sources significantly impacts the trajectory of a company’s development. At the same time, this provides an interesting premise for policy-making to influence the strategic clean energy technologies and cut the anthropogenic emissions.

Significance for Biofuels and Energy Efficiency Sectors

In terms of the sector variable, biofuels are the baseline sector in this model. From Table 2, the biofuels have the largest percentage of IPOs. In line with this observation, the model provides a negative coefficient for every other sector relative to the baseline (it means that the probability for any other sector is lower than for biofuels), with the only other significant sector being the energy efficiency sector (second place in total percentage of IPO exits).

Environmental, social, and economic implications. High statistical significance of the biofuels and energy efficiency to the outcome imply that these sectors are the most familiar and recognized by the public. Wider dissemination of these technologies will translate into environmental improvements through reductions in greenhouse gas emissions. The level of entrepreneurial risk has been reduced for these sectors, which provides an attractive entrepreneurial opportunity.

Significance for USA, China, and Canada

In the country variable, Canada is the baseline (3rd place in the IPO exits), after USA and China. Similarly, the coefficients for the latter are positive. The other countries do not have a significant relationship with the response variable.

Environmental, social and economic implications. US and Canada have one of the highest levels of greenhouse gasses per capita: 14.1 and 17.0 metric ton per capita, respectively (World Bank, 2015), while China is the largest gross emitter. The fact that these countries have achieved a high level of equity market maturity for clean technologies as well as public loans programs and venture capital shows a step in the right direction. Additional capital incentives for clean tech allow them to overcome the externalities in the cost of carbon (in terms of environmental and social costs).

Significance for the Top Ten Percent of Companies

Finally, the fact that the model for top 10^{th} percentile didn't show any significant variables implies that the null hypothesis was correct: the source of financing was not a statistically important determinant of the IPO exits among the top fundraising companies. I hypothesized that loans, structured debt, and Series A financing were important to reach the top tiers of financial maturity; however, among the top fundraising companies, the source of financing did not play any role in determining the IPO exit.

Environmental, social and economic implications. Once a company raised a significant amount of money, the source of funding didn't matter anymore, and market competitive mechanisms came in action. Once the cost of energy from renewable sources is competitive with the traditional energy sources (as is already the case in various cleantech subsectors), this market advantage will accelerate the dissemination of clean-energy solutions.

Recommendations for Further Research

The study offers significant insight into the general relationships between the financing structure of a startup and the IPO exits and revealed new questions for further research.

First, the model applied in this study could be improved by adding other response categories, such as "M&A exit," "Bankruptcy," and "Out of business." This would add more depth to understanding the relationship between company development and their financial structure. Methodologically, this study could be carried out by applying the nearest neighbor search classification methods.

Second, an interesting area for further research is adding annual information on changes in the startup status. The current data may be right censored in that some of the firms had not yet experienced an exit but could do so in the future. With this in mind, event history or survival analysis could be more appropriate, or perhaps a Poisson or negative binomial regression, with a firm-level random to treat time as discrete to analyze time-varying covariate effects, would work best.

Another interesting way to enhance the model would be to test for any possible interactions of the investment types. It is fair to assume that such interactions may have statistical significance.

Taking a step further, there is a broad field of topics to explore. What is the relationship between private and public capital or equity financing vs. non-dilutive capital? What are the significant financing sources in each cleantech sub-sector? What are the significant financial sources on a country level? Are there international spillovers of capital? What are the policy implications for the most risky and capital-intensive industries? How can the fundraising strategies for the startups in different sectors and countries be optimized? What are the implications for investors and policy makers?

Conclusions

Over the past 15 years, the early wave of clean innovation created a cleantech industry that is now reaching maturity in several of its subsectors (Jaffe, 2012), with a substantial number of cleantech companies going through IPOs and accessing public markets. I established a significant relationship between a possibility of an IPO exit and previous access to loans, structured debt, and Series A venture capital. This relationship is particularly significant for the biofuels and energy efficiency sectors. On the country level, I established a significant relationship with IPO exits and startups in USA, China, and Canada. At the same time, no statistical significance was established between the previous investment types and the IPO exits for the top fundraising startups.

References

Agresti, A. (2002). Categorical Data Analysis. Whiley - Interscience, John Whiley & Son, Inc. Publication. Retrieved from https://mathdept.iut.ac.ir/sites/mathdept.iut.ac.ir/files/AGRESTI.PDF

Bloomberg New Energy Finance. (2015a). New energy outlook 2015.

Bloomberg New Energy Finance. (2015b). Rebound in clean energy investment in 2014 beats expectations. Retrieved September 18, 2015, from http://about.bnef.com/press-releases/rebound-clean-energy-investment-2014-beats-expectations/

Bullard, N. (2014). Fossil fuel divestment : A $ 5 trillion challenge, (August), 1–19.

Caprotti, F. (2012). The cultural economy of cleantech: Environmental discourse and the emergence of a new technology sector. *Transactions of the Institute of British Geographers*, *37*(3), 370–385.

CEG, & Croatan Institute. (2014). What investors want: How to scale up demand for US clean energy and green bonds. Retrieved from http://www.cleanegroup.org/assets/Uploads/CEG-Croatan-What-Investors-Want.pdf

Cleantech Group. (2015). Connecting corporates with sustainable innovation. Retrieved September 7, 2015, from http://www.cleantech.com/

Criscuolo, C., & Menon, C. (2014). Environmental policies and risk finance in the green sector. *OECD Science, Technology and Industry Working Papers*, *2014*(01). http://doi.org/http://dx.doi.org/10.1787/5jz6wn918j37-en OECD

Frankfurt School-UNEP Centre, & BNEF. (2015). Global trends in renewable energy investment 2015.

Ghosh, S., & Nanda, R. (2010). Venture capital investment in the clean energy sector - Harvard Business School, 1–22.

Gort, M., & Klepper, S. (1982). Time paths in the diffusion of product innovations. *The Economic Journal*, *92*(367), 630–653. http://doi.org/10.2307/2232554

Hilbe, J. M. (2009). Logistic regression models - CRC Press Book. Retrieved from https://www.crcpress.com/Logistic-Regression-Models/Hilbe/9781420075755

Howell, S. T. (2014). Financing constraints as barriers to innovation: Evidence from R & D grants to energy startups.

Investopedia. (2015a). Growth fund definition. Retrieved September 16, 2015, from http://www.investopedia.com/terms/g/growthfund.asp

Investopedia. (2015b). Guaranteed loan definition. Retrieved September 16, 2015, from http://www.investopedia.com/terms/g/guaranteed-loan.asp

Investopedia. (2015c). Loan definition. Retrieved September 16, 2015, from http://www.investopedia.com/terms/l/loan.asp

Investopedia. (2015d). Project finance definition. Retrieved September 16, 2015, from http://www.investopedia.com/terms/p/projectfinance.asp

Investopedia. (2015e). Series A financing definition. Retrieved September 16, 2015, from http://www.investopedia.com/terms/s/seriesa.asp

Investopedia. (2015f). Series B financing definition. Retrieved September 16, 2015, from http://www.investopedia.com/terms/s/series-b-financing.asp

Investopedia. (2015g). Structured finance definition. Retrieved September 16, 2015, from http://www.investopedia.com/terms/s/structuredfinance.asp

International Renewable Energy Agency. (2015). Renewable power generation costs in 2014 : An overview. Retrieved from http://www.irena.org/ DocumentDownloads/Publications/Overview_Renewable Power Generation Costs in 2012.pdf

Jaffe, A. B. (2012). Technology policy and climate change, *3*(4), 1–15. http://doi.org/10.1142/S201000781250025X

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2006). An introduction to statistical learning with applications in R. *Design. 102*. http://doi.org/10.1016/j.peva.2007.06.006

de Jonge, E., & van der Loo, M. (2013). An introduction to data cleaning with R, 53.

Lester, R. K. (2009). America's energy innovation problem (and how to fix it). MIT-IPC-Energy Innovation Working Paper 09-007, (November), 53.

Lubin, D. A., & Esty, D. C. (2010). The sustainability imperative. Retrieved September 17, 2015, from https://hbr.org/2010/05/the-sustainability-imperative

Newell, R. G., Jaffe, A. B., Stavins, R. N., Newell, R. G., & Jaffe, A. B. (1998). The induced innovation hypothesis and energy-saving technological change, (Working Paper 6437).

NIST. (2005). Engineering statistics handbook. Retrieved from http://www.itl.nist.gov/div898/handbook/prc/section1/prc16.htm

Nordan, M. (2013). The state of cleantech venture capital: What lies ahead. Retrieved September 16, 2015, from http://mnordan.com/2013/03/27/the-state-of-cleantech-venture-capital-what-lies-ahead/

Ogden, P., Podesta, J., & Deutch, J. (2008). A new strategy to spur energy innovation. *Issues in Science and Technology*, *24*(2), 35–44.

Pachauri, R. K., Meyer, L., Van Ypersele, J.-P., Brinkman, S., Van Kesteren, L., Leprince-Ringuet, N., & Van Boxmeer, F. (2014). Assessment report 5, Climate Change 2014 Synthesis Report

Pagano, M., Panetta, F., & Zingales, L. (1998). Why do companies go public? An empitical analysis. *The Journal of Finance*.

Parad, M., & Cleantech Group. (2014). The global cleantech innovation index 2014.

PwC. (2015). Moneytree Report: Historical trend data. Retrieved September 21, 2015, from https://www.pwcmoneytree.com/HistoricTrends/CustomQueryHistoricTrend

Rogers, M. (2012). Energy = innovation : 10 disruptive technologies. McKinsey, 10–15.

SEIA. (2015). Solar energy. Retrieved September 20, 2015, from http://www.seia.org/about/solar-energy

UCLA. (2015). FAQ: How do I interpret odds ratios in logistic regression? Retrieved September 15, 2015, from http://www.ats.ucla.edu/stat/mult_pkg/faq/general/odds_ratio.htm

Verspagen, B. (2004). Patents, citations & innovations: A window on the knowledge economy. A.B. Jaffe, M. Trajtenberg (Eds.). MIT Press, (2002). Research Policy. http://doi.org/10.1016/j.respol.2004.06.002

Wang, J. (2013). Renewable energy projects: Tax-exempt and other tax-advantaged financing. Retrieved from http://corporatenl.eneco.nl/activiteiten/Duurzameprojecten/projecten-Nederland/Pages/Default.aspx

WilmerHale. (2014). 2014 IPO report.

World Bank. (2015). CO2 emissions (metric tons per capita). Retrieved September 29, 2015, from http://data.worldbank.org/indicator/EN.ATM.CO2E.PC

Appendix

Table 11. Binary investment variable model output: R regression summary.

```
Coefficients:
  Estimate Std. Error z value Pr(>|z|)
(Intercept)           4.374e+02  1.206e+02   3.627 0.000287 ***
  paidincap             2.176e-09  7.555e-10   2.880 0.003975 **
  sectorenergy storage -1.107e+00  6.811e-01  -1.626 0.103993
sectorenergyef       -1.142e+00  5.050e-01  -2.261 0.023763 *
  sectorhydro          -9.267e-01  1.200e+00  -0.772 0.439915
sectorsolar          -6.466e-01  4.966e-01  -1.302 0.192872
sectorwind           -7.932e-01  7.143e-01  -1.111 0.266763
countrycodeCN         1.770e+00  6.798e-01   2.604 0.009212 **
  countrycodeDE        -1.812e+01  9.488e+02  -0.019 0.984767
countrycodeFR        -7.627e-01  9.227e-01  -0.827 0.408484
countrycodeGB        -2.103e+00  1.141e+00  -1.844 0.065171 .
countrycodeIL        -7.271e-01  1.170e+00  -0.621 0.534338
countrycodeIT        -1.577e+01  2.217e+03  -0.007 0.994323
countrycodeNL        -1.773e+01  1.192e+03  -0.015 0.988129
countrycodeSE        -1.493e+00  1.356e+00  -1.102 0.270654
countrycodeUS        1.594e+00  5.860e-01  -2.720 0.006536 **
  year                 -2.188e-01  6.015e-02  -3.638 0.000275 ***
  num_rounds            1.068e-01  6.500e-02   1.644 0.100222
loan1                 1.739e+00  5.664e-01   3.071 0.002137 **
  strdebt1              1.082e+00  4.974e-01   2.175 0.029593 *
  seriesa1             1.112e+00  4.066e-01  -2.735 0.006232 **
  ---
  Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 402.23 on 1051 degrees of freedom

Residual deviance: 255.39 on 1031 degrees of freedom

AIC: 297.39

Number of Fisher Scoring iterations: 17

Table 12. Coefficient interpretation.

coefficients		odds	probabilities
(Intercept)	4.374094e+02	9.215203e+189	1.000000e+00
paidincap	2.176007e-09	1.000000e+00	5.000000e-01
sectorenergy storage	-1.107383e+00	3.304226e-01	2.483591e-01
sectorenergyef	-1.141710e+00	3.192724e-01	2.420065e-01
sectorhydro	-9.266536e-01	3.958763e-01	2.836041e-01
sectorsolar	-6.466280e-01	5.238091e-01	3.437498e-01
sectorwind	-7.932340e-01	4.523794e-01	3.114747e-01
countrycodeCN	1.770253e+00	5.872337e+00	8.544891e-01
countrycodeDE	-1.811602e+01	1.356168e-08	1.356168e-08
countrycodeFR	-7.626740e-01	4.664175e-01	3.180660e-01
countrycodeGB	-2.103410e+00	1.220395e-01	1.087658e-01
countrycodeIL	-7.270630e-01	4.833264e-01	3.258396e-01
countrycodeIT	-1.577408e+01	1.410606e-07	1.410606e-07
countrycodeNL	-1.773162e+01	1.991850e-08	1.991850e-08
countrycodeSE	-1.493164e+00	2.246608e-01	1.834473e-01
countrycodeUS	1.593768e+00	2.031587e-01	1.688544e-01
year	-2.188253e-01	8.034621e-01	4.455109e-01
num_rounds	1.068378e-01	1.112754e+00	5.266841e-01
loan1	1.739185e+00	5.692703e+00	8.505835e-01
strdebt1	1.082103e+00	2.950878e+00	7.468917e-01
seriesa1	1.112211e+00	3.288312e-01	2.474590e-01

Table 13. Continuous variable model output: R regression summary.

```
Coefficients:
  Estimate Std. Error z value Pr(>|z|)
(Intercept)          3.801e+02  1.190e+02   3.194 0.001402 **
  sectorenergy storage -1.269e+00  7.052e-01  -1.799 0.072032 .
sectorenergyef      -1.324e+00  5.053e-01  -2.620 0.008783 **
  sectorhydro         -9.599e-01  1.160e+00  -0.828 0.407877
sectorsolar         -4.653e-01  4.815e-01  -0.966 0.333913
sectorwind          -4.451e-01  6.642e-01  -0.670 0.502785
countrycodeCN        1.212e+00  6.302e-01   1.922 0.054563 .
countrycodeDE       -1.738e+01  1.014e+03  -0.017 0.986328
countrycodeFR       -1.254e+00  9.399e-01  -1.334 0.182208
countrycodeGB       -2.300e+00  1.123e+00  -2.049 0.040498 *
  countrycodeIL       -9.624e-01  1.143e+00  -0.842 0.399704
countrycodeIT       -1.691e+01  2.091e+03  -0.008 0.993546
countrycodeNL       -1.696e+01  1.300e+03  -0.013 0.989593
countrycodeSE       -7.998e-01  1.157e+00  -0.691 0.489555
countrycodeUS       -1.869e+00  5.491e-01  -3.405 0.000662 ***
  year                -1.903e-01  5.936e-02  -3.206 0.001344 **
  num_rounds           1.191e-01  5.663e-02   2.103 0.035443 *
  am_seed             -4.867e-07  4.010e-07  -1.214 0.224797
am_growtheq          5.061e-09  2.264e-09   2.235 0.025398 *
  am_structdeb         1.571e-08  6.842e-09   2.296 0.021704 *
  am_seriesa           3.024e-08  1.714e-08   1.764 0.077691 .
---
  Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 402.23 on 1051 degrees of freedom

Residual deviance: 265.97 on 1031 degrees of freedom

AIC: 307.97

Number of Fisher Scoring iterations: 17

Table 14. Coefficient interpretation.

coefficients		odds	probabilities
(Intercept)	3.800963e+02	1.184965e+165	1.000000e+00
sectorenergy storage	-1.268658e+00	2.812087e-01	2.194870e-01
sectorenergyef	-1.324112e+00	2.660391e-01	2.101350e-01
sectorhydro	-9.598711e-01	3.829422e-01	2.769040e-01
sectorsolar	-4.652573e-01	6.279735e-01	3.857394e-01
sectorwind	-4.451222e-01	6.407459e-01	3.905211e-01
countrycodeCN	1.211520e+00	3.358587e+00	7.705678e-01
countrycodeDE	-1.738280e+01	2.823237e-08	2.823237e-08
countrycodeFR	-1.253788e+00	2.854215e-01	2.220451e-01
countrycodeGB	-2.300260e+00	1.002328e-01	9.110146e-02
countrycodeIL	-9.623719e-01	3.819858e-01	2.764036e-01
countrycodeIT	-1.691370e+01	4.513099e-08	4.513099e-08
countrycodeNL	-1.696251e+01	4.298101e-08	4.298101e-08
countrycodeSE	-7.998186e-01	4.494105e-01	3.100643e-01
countrycodeUS	-1.869461e+00	1.542068e-01	1.336041e-01
year	-1.903374e-01	8.266801e-01	4.525588e-01
num_rounds	1.191098e-01	1.126494e+00	5.297423e-01
am_seed	-4.867347e-07	9.999995e-01	4.999999e-01
am_growtheq	5.060804e-09	1.000000e+00	5.000000e-01
am_structdeb	1.570564e-08	1.000000e+00	5.000000e-01
am_seriesa	3.023609e-08	1.000000e+00	5.000000e-01