ERASMUS UNIVERSITY ROTTERDAM Erasmus School of Economics

Bachelor Thesis [International Bachelor of Economics and Business Economics]

How going public affects innovation: evidence from China

Name student: Sungwai Lee

Student ID number: 524960

Supervisor: Bastian Ravesteijn

Second assessor: Coen van de Kraats

Date final version: 25 August 2022

Abstract

The decision of whether R&D is worth the investment and whether the company should go public are two decisions that many medium or large sized firms encounter. In this thesis, I explore the relationship between these two decisions, or more specifically attempt to manufacture a causal effect between firms going public and firm-level innovation in China. An OLS regression model is devised and the results show that the IPO process is indeed positively and significantly correlated with patent applications, the standard way of measuring innovation. However, statistical inference is rejected due to the amount of bias in the regression. Discussion and improvements suggest that the Chinese government can use this result to shape government policy to promote both innovation and going public.

1. **Introduction**

1.1 Introduction

When firms decide to go public, they make their shares available to purchase throughout an IPO process. Firms can opt to go public because of the increased reputation or prestige, stronger credibility, financial flexibility, and larger company size (*Bancel & Mittoo, 2009 and Pagano et. al, 1998*). The transformation from private to public by firms is one that contains many risks, and requires a large amount of consideration and careful decision making. Many factors, such as debt levels, amount of CEO ownership (*Latham, et. al, 2010*), and even local political factors (*Piotroski et. al, 2014*) are relevant when looking at why firms go through the IPO process, as well as how fast. Even though this transition is very closely examined by how it relates to firm performance (such as sales, revenue and growth) before and after IPO, there is not much known about the effects of going public on innovation (*Bernstein, 2015*). Furthermore, there is even less when considering the effects of the IPO process on innovation levels for Chinese firms specifically. This factor contributes heavily toward the scientific relevance for this thesis. The IPO process is initiated at a specific period of time during a firm's life cycle according to *Bernstein* (*2015*), and this thesis aims to help firms and governments to understand the direct and indirect effects of going public on their innovation levels. Even though this thesis will utilize the main findings of the paper by *Bernstein, S.* (*2015*), there is little other material that analyzes the impact of a firm's IPO on innovation. In addition, China has led the world in terms of patent filings since 2011, according to the World Intellectual Property Organization's (WIPO) report in 2020, and this thesis may provide some more understanding for the high levels of innovation in China specifically, such as whether firms going public will have an effect on these levels. The results from this thesis could be used to encourage more firms to go public, or for governments to support R&D, if a positive impact on innovation is found. How innovation benefits society can be seen from previous research based on the theory that innovation by firms facilitates economic growth, according to *Ahlstrom (2010), Akçomak & Ter Weel (2009), and Crosby (2000).* This provides some social relevance and reasoning for writing this paper. By gathering information from different previous papers and sources, this thesis can hopefully provide a larger picture for the direct and indirect effects of going public on a firms' innovation levels, and open up the field for more investigation as well.

1.2 Theoretical Framework

As mentioned before, much of this thesis will be based on the work of *Bernstein, S*. (*2015*). In theory, firms that choose to go public should not undergo any direct changes to their innovation activity. Firms that decide to enter the public equity market may already take part in innovation processes to a certain extent. Firms in industries that have already entered the later stages of the industry life cycle are likely to have different strategies pursuing innovation. The industry life cycle is defined by five stages through a graph of time against the number of firms in the industry (*Keppler, 1997*). The later stages of the industry life cycle follows from a period when there is low or zero entry, and firms are less likely to increase their innovation levels as a result of declining entry into the industry and the emergence of new market equilibria. Dominant design theory describes the stage of an industry in which innovation drastically declines and the optimal and dominant production of the main product is revealed and mastered (*Suarez and Utterback*, *1995*). Firms in the industry have all or mostly obtained the information on how to produce the dominant design product, as the firms that do not produce this cannot survive in the industry. This changes the underlying competition strategies for firms in the industry, where firms will compete against each other purely on minimizing production costs or lowering prices. Industries in which the main product of the industry has achieved a dominant design are also unlikely to increase their innovation levels as there is much less emphasis on R&D expenditures. In addition, small, high growth firms such as gazelles (firms that exhibit a 20% increase in revenue over at least a four year period) that may have exploited the substantial innovation levels of an existing industry to its advantage are less willing to contribute to the development of new technologies, as their emphasis is mainly on firm growth. All of the above maintain the stance that regardless of whether a firm lists the shares of its corporation to the public or decides to remain private, it should not have an effect on the level of innovation the firm undergoes.

However, the decision to follow through with the IPO process for firms may stimulate their innovation levels through specific indirect effects. As mentioned before, firms can choose to go public to achieve stronger prestige and attract more attention to their brand name. This creates an underlying effect of generating more sales because of their increased attention, at the cost of part of the company's ownership. Overall, the firm can achieve the motive of raising capital, such as the acquiring of cash or lowering of firm debt. In doing so, firms are able to access increased and new human capital, such as more prestigious scientists, innovators, managers or leaders, as well as the acquisition of productivity or R&D enhancing technologies (*Bernstein, 2015*). This changes the strategies that firms pursue in terms of innovation levels (*Bernstein, 2015*). Furthermore, changes in ownership dilution and the structure of the firm may affect ongoing and future projects, which weakens the incentive for inventors at the firm to innovate (*Bernstein, 2015*). Some reasons for this is that inventors become more wealthy after their claims to their inventions, and that managers prefer to adopt existing but efficient technologies that are unambiguous to the market, and therefore less likely to fail. All of these factors suggest that going public does in fact have an effect on the innovation levels of a firm.

One important observation that must be taken into consideration is that there is a natural selection bias in this thesis; that is, firms are choosing by themselves whether or not to go public. This is solved in the paper by *Bernstein,* (2015) by examining firms that either complete or withdraw their IPO filing, instead of taking all firms into account. This method, however, does not remove the selection bias, as it is not random whether firms choose to complete or withdraw the IPO filing. These issues will be addressed later on in the thesis.

In order to measure innovation, many research papers decide to use the indicator of how many patents are granted to the firm (*Archibugi, 1992 and Dang, 2015 and Nagaoka, 2010)*. However, the paper by *Bernstein* (2015) reflects the problem that patent counts are unable to distinguish between breakthrough, industry changing inventions and minor discoveries, which can both make a case for a patent filing. Thus, to realize the originality of the patent, or how impactful the discovery can be, the number of citations that a patent receives after being approved can be used to resolve this problem.

1.3 Topic justification

Before looking at the effect of going public on innovation, some background knowledge and topic motivation will be introduced. As mentioned before, innovation is one of the core strategies for promoting economic growth (*Solow, 1957*). However, there are multiple reasons as to why innovation is weakly incentivized. This may be due to the free rider problem: if a firm invests heavily into R&D and information is leaked or released to the public, others can benefit from inventions or innovations without cost. The most common methods to counteract this problem is usually through the form of intellectual property rights (IPR) or patents, which this paper will study more about. Another reason why innovation is not incentivised for firms is because research is usually cumulative, in the sense that people often build on the work of others. Firms may be the leader of innovation in the field, yet other firms can then use this newly invented information and improve upon the idea. This causes the original firm's idea to become less prevalent and obsolete, while other firms can take advantage of these ideas and build their products upon the inventions. Furthermore, firms may innovate ideas and products so advanced that they would even prefer to keep their information about their breakthroughs secret, such that competitors are kept in the dark about their advancement. However, if the innovative products are kept secret, this means that the firms cannot make profits or benefit financially from the innovation. In fact, this likely discourages the firm to invest time and resources into developing innovative but unprofitable products and technology; they would much rather have other firms develop these resources and build new ideas and improvements upon it. Finally, there also exists situations where patents cannot prevent infringement, such as compulsory licensing. This is a situation where firms in other countries are not required to follow the original firm's patenting rights. This is usually to help developing countries or in emergency crises, such as medicinal development for diseases across the world.

Patents are a complicated method to increase the incentive for firms to innovate. Firstly, there is a lack of evidence that points towards strong patent systems increasing R&D (*Williams, 2017*). Additionally, the design of the patents itself can be a problem; the patent must be sufficiently long over a period of time, but also sufficiently broad such that other competitors cannot find ways to skirt around the rules of the patent. As mentioned before, sometimes firms in certain industries prefer to keep inventions or scientific breakthroughs a secret, because of the possibility of gaining market advantages by implementing the invention in certain situations. One way to promote patenting would be the use of patent pools, where two or more parties pool their technology and license them as a package. This can help to reduce transaction costs, as well as integrate complementary technologies, avoiding infringement litigation and more certainty of knowledge access. All these benefits point toward that in theory, patent pools should be an incentive for innovation. In contrast, firms that join together in the pool (who usually have a large number of patents before joining the pool) have a reduced number of patents, and firms outside the pool have increased patents (*Lampe et.al, 2010*) after the pool is formed. There is evidence that forming a patent pool may discourage firms with a large number of patents before joining the pool, which are inherently the larger and more resourceful firms.

As for region justification, I decided to choose China because it has an extremely high growth rate when it comes to innovation and patenting. Figure 1 below shows the growth rate of patent applications per country from 2010 to 2020. This data was collected from the World Data Bank, under the variable “resident patent applications”. The sample size consists of countries in the G20 neighboring or in a close vicinity of the country of interest, China as well as the UK and USA as standard benchmarks. The data collected consisted of the absolute values of patent applications from the selected countries, but absolute values are inaccurate and incomparable due to the vast population of China heavily impacting the magnitude of the numbers. Therefore, a simple growth rate for each country was computed, with 2010 being the base year, and the year 2011 stands for the rate of growth in patent applications from 2010 to 2011.

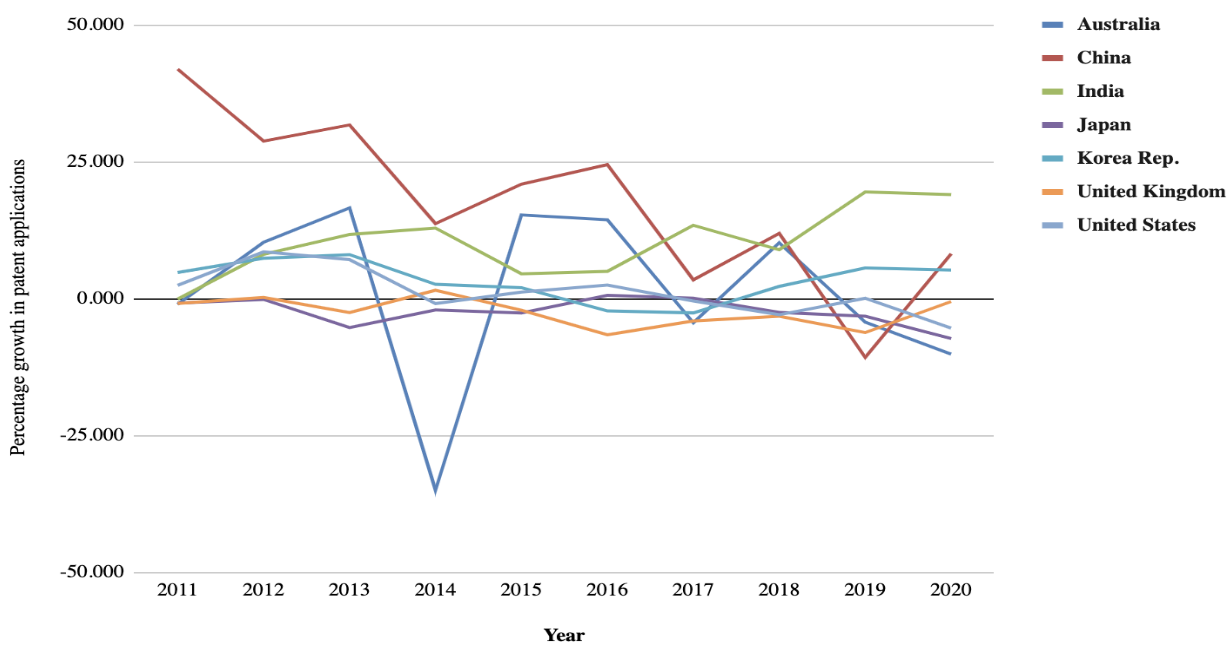

Figure 1: Percentage growth of patent applications per year for selected countries in G20

**Source: World Data Bank 2022*

Figure 1 shows the percentage growth of patent applications each year for the country of interest, China, as well as some other countries for comparison. As can be seen, China has consistently a higher growth rate than most, if not all of the countries for comparison. Although the line shows a decreasing trend, it should be noted that growth remains positive since 2010 until 2019, meaning that the magnitude of total patent applications is always increasing. China's percentage growth of patent applications is also one of the countries that fluctuate strongly, which prompts investigation into the cause behind this fluctuation.

2. Data and Methodology

In this section I will discuss the data used in this thesis, as well as the method in which I will attempt to establish a causal relationship between going public and firm innovation. Firstly, it must be mentioned that data collection was severely limited due to the relatively smaller sample size (compared to the sample size of Bernstein's paper) due to the area restraint of the data. IPO data for Chinese firms are not readily accessible. Therefore, this sample size will consist of 26 annual aggregate observations, collected from the World Data Bank. Each observation will be an annual aggregate year, starting from 1995 up to 2020, with the corresponding number of total listed companies that year and the number of patent applications in said year.

For this thesis, a multivariate time series regression model will be used to construct the model in order to find any difference in the impact of more Chinese firms going public on innovation. Therefore, the dependent variable of the regression will be the innovation level, measured by the number of patent applications from each year. The variable of interest will be the total number of listed companies for each year. The coefficient of this model can be interpreted as the following: one more firm going public and being listed on the market will lead to a increase or decrease in β patents. The regression will be run in Stata, and will consist of heteroskedastic standard errors.
Due to the nature of the data, an autocorrelation test is run in Stata using the Durbin-Watson test, which resulted in a value of 0.686. Comparing this to the lower bound value in the Durbin-Watson statistic table, there is enough evidence to say that the data is positively autocorrelated. Some control variables, variables that are correlated with both the dependent

and independent variable, are further added to reduce the bias of the regression. Firstly, the control variable of population is added to the model. This variable is defined as the population of people aged 15-64 as a percentage of total population, retrieved from the World Data Bank. Another control variable added to the model will be the unemployment rate, defined as the total unemployment rate out of the total labour force. Some justification for the addition of this variable is because of its shown correlation with innovation (*De Elejalde et.al., 2015*) as well as going public (*Johannson, A et.al., 2017*). With these variables, the OLS regression method will be conducted to find any causal relationship between the variable of interest and the independent variable, the process of going public. Therefore, the regression equation for the model will be formulated as follows:

$$Y = \alpha + \beta IPO + \gamma X_1 + \delta X_2 + \varepsilon$$

$Y = dependent\ variable\ (count\ of\ annual\ patent\ applications)$

$\alpha = constant$

$IPO = count\ of\ annual\ listed\ companies$

$\beta = IPO\ coefficient$

$X_1, X_2 = control\ variables,\ population\ and\ unemployment$

$\gamma,\ \delta = control\ variable\ coefficients$

$\varepsilon = error\ term$

Several hypotheses will now be drawn in order to outline the expectations of the model. Firstly, the relationship between undertaking the IPO process and the increase in patents is expected to be positive. Some evidence, such as the recognition that going public can change a firm's strategy about innovation (*Bernstein, S., 2015*), can be used to support this thesis. Even though the variables in this model will be more different, and the analysis is more on a country level basis rather than a firm level basis as the one used by Bernstein due to the variables used in the model, the hypothesis will be expecting a positive correlation and, if applicable, causal effect between going public and innovation. In regard to the other control variables, as mentioned before, they were chosen due to their assumed correlation with the IPO variable and the error term. The population variable is expected, as a hypothesis, to have a positive relationship with innovation. Intuitively, this makes sense as more people are able to have a higher chance of contributing towards making innovative changes. Larger population gives a larger probability to generating more individuals that are able to develop and engineer more breakthrough ideas. Crucially, this indicates that the sign of the omitted variable bias will be negative, due to this positive correlation. Similarly, more unemployment will intuitively lead to a lower level of innovation, as there are less people to generate innovative ideas. A negative relationship between unemployment and innovation is expected for the regression, and the sign of this omitted variable bias may result as positive.

First hypothesis*: there is a positive relationship between a company going public and the number of annual patent applications.*

Second hypothesis*: a positive covariance is expected between the population variable, defined as the total population aged 15-64, and the annual patent application count.*

Third hypothesis: *for the relationship between the unemployment variable, defined as a percentage out of the total labour force, a negative covariance is predicted*

Table 1 below gives some summary statistics for the data in this sample. The mean, standard deviation, minimum and maximum values for each of the four variables are given. The standard deviation of the independent and dependent variable immediately stands out to be almost as large as, or larger than, the mean itself. This is due to the larger change over the years in the number of patent applications as well as the number of public companies in China, measured annually. The control variables in the regression equation, population and unemployment, are relatively stable over the last 26 years, with the ranges for the observations also relatively close together. This data will now be used to estimate coefficients and determine the relationship between the regressors.

Variable	Mean	Standard Deviation	Minimum	Maximum
listed companies	1928.3	1060.8	323	4154
patent applications	431745.8	499022.1	10011	1393815
population	70.95%	2.215	66.84	73.27
unemployment	3.85%	0.556	2.9	5.15

Table 1: summary statistics of regressors (number of observations: 26)

3. Results and analysis

3.1 Results

Table 2 demonstrates the regression results of the OLS regression multivariate analysis. I include two models, one with and one without the control variables devised earlier, in order to control for selection bias. Both columns (1) and (2) have the dependent variable as the number of annual patent applications, while column (1) only the listed companies variable (measured by annual listed companies) and column (2) includes the control variables that are added to control for selection bias to some extent.

As can be seen from the coefficient of model (1), there is a positive and significant relationship between the number of listed companies and the number of patent applications in this model. More precisely, the coefficient represents a 451.8 average increase in the count of patent applications when a company becomes listed (i.e going public), and this effect is significant at the $p < 0.001$ level. The constant of the first model is negative and significant, meaning the null hypothesis that the constant is equal to zero can be rejected. The sign and magnitude of the constant, however, is more interesting. The constant can be interpreted as the number of patent applications when there are no listed companies. However, to reach a positive number of patent applications, the number of listed firms is required to be at least 973 listed companies, which is larger than several of the observations. This is a potential issue that will be discussed in the later sections.

Model (2) adds both control variables in population and unemployment to the regression model, in order to check for their significance and reduce some selection bias in the model. Primarily, the coefficient for the independent variable has changed, and has seen an 80 unit increase in magnitude. Therefore, a company going public is likely to create an average increase of 530.9 patent applications. This coefficient is also significant at the $p < 0.001$ level. Meanwhile, the population coefficient is negative and significant at the $p < 0.05$ level. This directly contrasts the second hypothesis generated earlier, where population was predicted to have a positive correlation with patent applications. The potential reasons and explanations for this will be further discussed in the analysis. Finally, the coefficient for the unemployment variable is also negative and significant at the $p < 0.05$ level, and it is consistent with the third hypothesis discussed earlier.

	(1)	(2)
	patent applications	patent applications
listed companies	451.8***	530.9***
	(26.75)	(26.69)
population		-30617.3*
		(13269.3)
unemployment		-129783.0*
		(58839.6)
constant	-439471.1***	2078219.4*
	(58602.9)	(802322.5)
N	26	23
R^2	0.922	0.964
Adjusted R^2	0.919	0.959

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2: relationship between IPO and patent applications

3.2 Selection bias

This section will be dedicated to discussing the selection bias in this regression. Firstly, selection bias in a regression context can also be referred to as omitted variables bias (OVB), which is one of the key assumptions for the OLS regression method if a causal effect is to be estimated. Essentially, to ensure the validity of statistical inference in an OLS regression, all the variables that are correlated with both the variable of interest and the IPO variable must be included within the regression to be controlled for. However, due to data restrictions, not all the variables that are correlated with both applying for patents and going public are included in the regression. For example, data such as firm age, industry cycles and even how other countries are performing in regard to economic performance are included in the regression. By not including these variables (and likely other unobserved variables), the assumption that the error term is uncorrelated with the regressors is violated.

Furthermore, when collecting and evaluating the data, there is an inherent selection bias when it comes to the regressors; that is, firms themselves are choosing whether or not to undertake the IPO process, and they are not chosen randomly. In fact, all the observations in the listed companies variable are fundamentally biased, as all of them had applied for the IPO process willingly. Therefore, this is a violation of the conditional independence assumption, which regulates the internal validity of the regression, rendering a causal effect even more unlikely as a result.

3.3 Interpretation and discussion of results

Some more in depth discussion about the results of the coefficients will be discussed in this section. For both models, the relationship between the constants and the variable of interest, patent applications, must be discussed. More specifically, the model estimates an average of 451.8 extra patent applications when a company lists itself publicly, which seems highly unlikely. However, this may be due to the format of the variables. The patent applications variable looks at the total number of applications submitted for the entire country of China annually, and this includes companies that are not listed publicly as well as those that are listed publicly. Hence, not all and most likely only a small part of the increase in the estimated coefficient is due to the listing of a firm onto the market. This may also be the reason as to why the first model has a constant that is negative; patent applications are not entirely dependent on companies going public.

Additionally, the second hypothesis devised in section 2 mentioned that the variable population was expected to have a positive correlation, if not significant, with the count of patent applications. However, the estimated coefficient resulted in a negative sign, thus demonstrating a negative relationship between an increase in 1% of the total population of China and the number of annual patent applications. One reason for this, when looking closer into the details of the model, may be caused by the large magnitude of the constant. Unlike the first model, model (2) estimates a significantly larger constant for the regression equation, a number exceeding two million. However, also unlike the first model, there are two other variables controlled for in the regression. Since the number of patent applications is likely different (although not very far off) to the estimate in model (1) when there are no listed companies, the negative sign of the population variable (as well as the unemployment variable) is to compensate for the significant magnitude of the constant, in order to generate a result similar to the estimated constant in model (1). As a result, the sign, as well as the magnitude, of the population coefficient is required to be large and negative. This applies in the same way for the unemployment variable, although in contrast to the population variable, estimated the coefficient in a way that fulfilled the third hypothesis. However, the magnitude of the coefficient is significantly larger than compared to the population coefficient. This can be explained by the underlying difference in the definition of both variables. The unemployment variable measured unemployment as a percentage of the total labour force, and the coefficient in model (2) for unemployment represents the effect of a 1% increase in unemployment on the total number of patents. An increase in 1% of unemployment is a relatively significant effect overall, considering that the range of total unemployment over the last 26 years is 2.15%.

4. Discussion and conclusion

4.1 Model, data and method discussion

This section will now address some of the problems regarding the selected regression model and the design of the experiment. Firstly, as mentioned before in section 3.2, there is an inherent bias in the regression. This is due to the lack of available data that matches with the main regressor data, listed companies and patent applications. Some methods can be devised to solve the selection bias, which will be discussed more in detail in the improvements section.

As such, a causal effect cannot be estimated using these regression models, even though the coefficient of the independent variable is significant at the $p < 0.001$ level. This model can only provide the information that the number of public companies is positively correlated with the number of annual patent applications, but conclusions cannot be drawn about the number of applications increasing being caused by an increase in companies undergoing the IPO process.

The main restricting factor throughout this thesis was mainly down to the lack of accessible and complete data. I had originally planned on running the regression model on a firm-specific level, similar to the method used in Berstein’s paper. However, the limited data, and especially specific data on Chinese firms, was not readily available, and thus the attention was switched to analyzing the effects of going public on innovation on a country level. There are some issues with this switch. Both innovation and the IPO process are typically associated as firm-level decisions, and analyzing the results of a regression based on a country level can lower the external validity of the regression, i.e how applicable the results are in the real world. Ultimately, the IPO decision and firm innovation is a decision that is made by firms at the company level, rather than a country level and which is not controlled by the government or government policy. The results generated in this regression model can rather be used as some information provided for the government of China, on how their domestic firms are innovating and how much of the innovation is down to the IPO process. As mentioned before, innovation is also not entirely dependent on public and listed firms, as private firms are also filing for patents parallel to public firms. Additionally, as a result of the regression being modeled on a country level instead of a firm level analysis, the control variables are chosen differently and less adequately. For example, when doing a firm level analysis on the effect of firms going public on overall innovation, variables such as firm age, patent application year and other within-firm characteristics are controlled for, variables that are correlated with the IPO variable and the error term. The inclusion of these variables allows for a more specific and in-depth picture of the actual effects of an IPO process on the innovation levels of a firm, and the conclusions drawn from that type of data may be useful in providing firms with an argument to go public.

4.2 Interpretation of results

The results from the OLS regression are unconvincing when trying to estimate statistical inference, due to the large amount of bias and non-ideal sample size and composition. It also appears inapplicable on a firm-level basis, as the variables and data are all constructed on a country basis. However, this means that the country, or rather the government of China, can partly use this information as reference to promote the switch for firms from private to public. Ultimately, innovation can create a deadweight loss to society due to the nature of patents, but in return firms' intellectual property are protected. This is because patents create a monopoly system for the company that files the patent. The government must then weigh the deadweight loss created by the patent system and measure it with the benefits that come from firms investing in R&D and therefore innovating. An example of this could be promoting medical firms to innovate in order to engineer medicine and vaccines. The information that this regression model offers can therefore be of some, if not limited, use, to aid the government in understanding the R&D investing and IPO situations. Government policy can then be altered in order to incentivize innovation, as well as the decision to go public, and potentially become more lenient with previous regulations. There has been existing evidence that government policy promoting innovation can have an effect on firm level innovation, depending on factors such as a firm's willingness, capacity and opportunity to change (*Patanakul, P., & Pinto, J. K., 2014*). Furthermore, firms can be willing to go public due to changes in government and political stances and policies, as they can reap large benefits during the IPO process (*Francis, B. B. et.al., 2009*). Due to the results of the regression model devised in this thesis, even though it results are lacking in validity, it can help shape certain government policies regarding the regulations of both innovating and the process of listing a company as public

4.3 Improvements

Several improvements are suggested in this section to provide some insight for further research, as well as ways that this regression model could have been modified for more meaningful results. As mentioned before, the lack of other control variables significantly hinder the validity of the regression, and including additional control variables into the regression model could largely reduce omitted variable bias, thus providing a more convincing argument for statistical inference. Since omitted variable bias is most likely never completely removable, due to unobservable variables correlated with the error term, the best that can be done would be to reduce it to the greatest extent and increase the credibility of the model. One detail that should also be improved on is the measurement of the dependent variable. In section 1.2 I argued that the number of patent citations, rather than looking at patents in general, should be used in order to provide a more accurate account of innovation. However, this data was unavailable, meaning that the external validity of the regression is also slightly compromised.

Overall, the design of the experiment on a country level basis rather than looking at Chinese firms was also a problem. The results generated provided more of an insight for governments to adjust government policy, rather than aid firms in making a decision whether or not to go public. Therefore, an improvement in the design of this experiment, making the assumption that data on public Chinese firms and that matching data for the control variables is ready and available, could be to switch to exploring the causal relationships within Chinese firms. With this switch, other control variables mentioned before such as firm age or patent application year could be used to control for omitted variable bias. Firm age is likely correlated with innovation as well as the decision to go public, as younger firms are likely to undergo different strategies regarding innovation (*Coad, A. et. al., 2016*). Firm age also plays a role in determining when, as well as the quality of performance, a firm decides to list its shares publicly, which makes it a crucial control variable as part of the improved regression. Another control variable suitable for improving the design of the regression is the year that patents were filed. Certain periods, such as the financial regression in 2008, represent an unsuitable and unideal economic environment for firms to turn public, or invest in firm innovation. Therefore, the addition of a dummy variable that equals one when the application year is a year of economic downturn and zero when the year is not makes the results of the regression more convincing if it is controlled for. Regarding the inherent selection bias of firms choosing by themselves to go public, a solution is also provided in the paper by *Bernstein, S.* If given the correct and suitable data, firms could be divided into two firms that filed for the IPO process and followed through with going public, or firms that withdrew from the filing (*Bernstein, S., 2015*). A dummy variable that gives a value of one when firms completed the process and zero when they withdrew would largely reduce the selection bias that existed in the regression above, thus guaranteeing higher internal validity as well.

Another way to improve this experiment could be the addition of an instrumental variable, or IV. The instrumental variables method, unlike the OLS regression method, does not require all observed variables to be controlled for. The conditions for IV to generate a causal effect include an assumption that the instrumental variable is independent from all other control variables that could have an effect on the independent variable and variable of interest. Even though this assumption can never be verified, informal checks on the correlation between the IV and observed variables can be run to check on the independence assumption. Regardless, the IV method is one that can provide significantly large internal validity, at the expense of more external validity, due to its estimation of the local average treatment effect rather than the average treatment effect. The addition of an instrument variable as a second stage to the regression could improve the model's internal validity, as it disregards all the omitted variable bias in the regression.

4.4 Conclusion and further research

This thesis attempted to provide empirical evidence for the relationship and effects between going public on innovation levels in China. The OLS regression method that was used provided some insight, with two of the three hypotheses generated in the methodology section proven correct, as well as providing significant coefficients in the estimation. However, due to the omitted variable bias and selection bias in the regression, no causal effect can be estimated. Recommendations on how to improve the regression were given in the improvements section, including methods to change the experiment focus from a country level to a within-firm level. If those recommendations are achieved and developed, the existence of statistical inference on evidence of the relationship between undergoing IPO and innovation could be vital for firms struggling with the decision to go public, especially regarding Chinese firms. Examples of further research could be an investigation into other countries' firms and experiment with whether firms in other countries behave as such. Those experiments should focus on establishing high internal validity, as the regression goal should be to establish a firm causal relation between the selected country's company IPO decision and their innovation levels.

References

Ahlstrom, D. (2010). Innovation and growth: How business contributes to society. *Academy of Management Perspectives*, 24(3), 11-24.

Akçomak, I. S., & Ter Weel, B. (2009). Social capital, innovation and growth: Evidence from Europe. *European Economic Review*, *53*(5), 544-567.

Archibugi, D. (1992). Patenting as an indicator of technological innovation: a review. *Science and Public Policy*, *19*(6), 357-368.

Bancel, F., & Mittoo, U. R. (2009). Why do European firms go public?. *European Financial Management*, *15*(4), 844-884.

Bernstein, S. (2015). Does going public affect innovation?. *The Journal of Finance*, *70*(4), 1365-1403.

Bleda, M., Morrison, K., & Rigby, J. (2013). The role and importance of gazelles and other growth firms for innovation and competitiveness. *Innovation Policy Challenges for the 21st Century*, *27*, 110.

Chen, Y., & Puttitanun, T. (2005). Intellectual property rights and innovation in developing countries. *Journal of Development Economics*, *78*(2), 474-493.

Coad, A., Segarra, A., & Teruel, M. (2016). Innovation and firm growth: does firm age play a role?. *Research Policy*, 45(2), 387-400.

Crosby, M. (2000). Patents, innovation and growth. *Economic Record*, *76*(234), 255-262.

Dang, J., & Motohashi, K. (2015). Patent statistics: A good indicator for innovation in China? Patent subsidy program impacts on patent quality. *China Economic Review*, *35*, 137-155.

De Elejalde, R., Giuliodori, D., & Stucchi, R. (2015). Employment and innovation: Firm-level evidence from Argentina. *Emerging Markets Finance and Trade*, 51(1), 27-47.

Francis, B. B., Hasan, I., & Sun, X. (2009). Political connections and the process of going public: Evidence from China. *Journal of International Money and Finance,* 28(4), 696-719.

Johansson, A. C., Luo, D., Rickne, J., & Zheng, W. (2017). Government intervention in the capital allocation process: Excess employment as an IPO selection rule in China. *China Economic Review*, 44, 271-281.

Klepper, S. (1997). Industry life cycles. *Industrial and Corporate Change*, *6*(1), 145-182.

Lampe, R., & Moser, P. (2010). Do patent pools encourage innovation? Evidence from the nineteenth-century sewing machine industry. *The Journal of Economic History*, 70(4), 898-920.

Latham, S., & Braun, M. R. (2010). To IPO or not to IPO: Risks, uncertainty and the decision to go public. *British Journal of Management,* 21(3), 666-683.

Leach, M., Rockström, J., Raskin, P., Scoones, I., Stirling, A. C., Smith, A., Thompson, J., Millstone, E., Ely, A., Arond, E., Folke, C., & Olsson, P. (2012). Transforming innovation for sustainability. *Ecology and Society*, 17(2).

Nagaoka, S., Motohashi, K., & Goto, A. (2010). Patent statistics as an innovation indicator. *Handbook of the Economics of Innovation*, 2, 1083-1127.

Pagano, M., Panetta, F., & Zingales, L. (1998). Why do companies go public? An empirical analysis. *The Journal of Finance*, *53*(1), 27-64.

Patanakul, P., & Pinto, J. K. (2014). Examining the roles of government policy on innovation. *The Journal of High Technology Management Research,* 25(2), 97-107.

Piotroski, J. D., & Zhang, T. (2014). Politicians and the IPO decision: The impact of impending political promotions on IPO activity in China. *Journal of Financial Economics,* 111(1), 111-136.

Solow, R. M. (1957). Technical change and the aggregate production function. *The Review of Economics and Statistics,* 312-320.

Suarez, F. F., & Utterback, J. M. (1995). Dominant designs and the survival of firms. *Strategic Management Journal*, *16*(6), 415-430.

The World Bank, World Development Indicators (2012).

Williams, H. L. (2017). How do patents affect research investments?.