HEC MONTRÉAL

Does post-privatization risk taking translate into more innovation?

par

Axar Patel

Pascale Valery
HEC Montréal
Directrice de recherche

Sciences de la gestion
(Spécialisation Finance)

Mémoire présenté en vue de l'obtention
du grade de maîtrise ès sciences
(M. Sc.)

August 2021

Abstract

Using a unique database of 246 newly privatized firms from 19 countries, we investigate the impact of corporate risk-taking behaviour on corporate innovation. We find strong and robust evidence that corporate risk-taking is positively related to corporate innovation. Additionally, we also investigate the impact of collaboration and state ownership on corporate innovation. Our results suggest that collaboration and state ownership have a negative impact on corporate innovation.

Keywords

Privatization, innovation, risk, collaboration, ownership structure

Contents

List of Tables

Introduction

Privatization has important economic implications, particularly as it was a widespread practice in the 1980s and 1990s. The effects of privatization can be mixed, and some firms go back and forth in terms of ownership structure between being private and state owned. One such example is Air Canada which started as a private organization, then became a crown corporation, and finally in 1988 privatized. Identifying the benefits and weaknesses of privatization could help us understand why some firm switch ownership structure multiple times, or why firms privatize in the first place.

The issue of privatization is also one that may be interesting to governments. State owned enterprises (SOE) may lead to social and economic growth and stability. It may also create jobs, as their main goal may not be profit driven but socially and politically motivated. Furthermore, in the pursuit of economic growth, it may be that SOEs have better innovation practices as it is well known that innovation drives growth (Ahlstrom (2010)). Perhaps private firms have better innovation practices as they seek profit growth and are under more pressure than SOEs.

Therefore, research on privatization's effects on innovation is an economically important topic as it may dictate, or at least enlighten governments, private economic agents, and managers of both SOEs and non-SOEs in their decisions. It can help private agents make decisions on whether to buy an SOE, it can help governments decide whether to privatize an SOE or buy a private enterprise, it can guide decision makers in the process of privatizing. Finally, it can help leverage the benefits and strength of each type of firm while mitigating their weaknesses.

The effects of various aspects such as risk, collaboration and ownership on innovation activities of the privatized firm is unexplored. As an extension to Boubakri et al. (2013) paper, which explores the effect of ownership structure on risk taking behaviour of the firm, we will explore the effects of risk on innovation of privatized firms. Further, we will also explore the effects of collaboration and state ownership on innovation of the firm. Specific questions of interests are:

1. Does post-privatization risk-taking translate into more innovation?
2. How does collaboration impact the innovation activities of the firm?
3. How does state ownership relate to the innovation activities of the firm?

Contributions

We made a few contributions to the existing literature with this study. With data on innovation gathered from Derwent Innovation, we completed the database on privatisation. We investigated the impact of risk-taking on innovation. We also looked into the impact of collaboration on innovation activities by a firm. We also conducted an analysis as an extension to Boubakri et al. (2013) to see if there was a link between state ownership and innovation.

Results

A negative binomial regression was used to test the hypothesis that privatized companies that are more risk taking should be expected to engage in more innovation activities. We found that risk taking has a positive effect on the innovation activities of the firm. These results were corroborated through additional robustness tests. We also tested the hypothesis that more collaboration should lead to more innovation activities. Contrary to our hypothesis, we found that collaboration and innovation are negatively related. Using the same model, we also tested the hypothesis that state ownership is negatively related to the innovation activities of a firm. We found that state ownership is negatively correlated with the innovation activities of a firm.

Literature review

The effects of privatization on different aspects of the firm such as efficiency (Okten and Arin (2006)), innovation (Somé et al. (2021)), market liquidity and pattern of share ownership (Boutchkova and Megginson (2000)) have been studied by multiple authors who approach the issue at hand in different ways. Somé et al. (2021) hypothesize that private firms will narrow their innovations. Also, they believe privatization will increase the geographical diversification of talent sourcing. This last point may help in driving innovations as firms gain access to a much broader source of talent, best practices in innovation management, and differing views on what diversification policy should be for the firm. The authors point out that public enterprises have proved to be highly inefficient because they pursue strategies such as excess unemployment that satisfy the political agenda of the politicians who control them. So, after the divestiture of the government, the newly privatized firms operate with greater efficiency, better management talent, better corporate governance, efficient capital allocation and innovation patterns.

Newly privatized firms are likely to change their innovation patterns. Two primary measures for gauging innovation patterns of firms are patenting activities such as citations count received on the patents filed by the company and cross-country collaboration for technological innovation. Private firms may patent more or less than state owned enterprises. The pressure for short term results may affect the technological focus of the company as private firms invest their resources in the technologies with highest potential commercial success. Regardless of the level of spending on research and development, newly privatized firms realign their focus only on business goals rather than broader national interest. As a result, newly privatized firms focus only on most promising technologies and reduces the technological diversification. There is a consensus in the existing literature on privatization that private firms outperform State Owned Enterprise (Boubakri and Cosset (1998)) and that the main perceived benefits of privatizations are increased efficiency (Okten and Arin (2006)), and growth of the stock markets (Boutchkova and Megginson (2000)). Somé et al. (2021) argue that this is due to diverging interests in SOEs and private firms. Indeed, they attribute inefficiency to interests that are politically and socially motivated. Perhaps firms act in a way that would give the current government an edge over the next elections. Also, SOEs are not faced with the pressures of short-term performance results as they are state backed. Also, private firms are less subject to political constraints.

The paper by Tan et al. (2020) is very interesting as the authors directly relate privatisation, albeit partial, to corporate innovation. The authors find that partial privatization of China's state-owned firms have a positive effect on innovation. It is important to understand the impact of partial privatization because privatization transactions begin with a private sale of equity. The paper explains that privatization can spur innovation due to several reasons. It mitigates the agency problem between government agents and private shareholders. A primary concern for state government agents is of resource allocation. Better alignment by partial privatization could lead to more efficient resource allocation. Also, the partial privatization brings more information about the state-owned enterprises which reduces the information asymmetry which might be used by shareholders to monitor managers and allow managers to make more informed corporate investments such as the technological innovation. On the other hand, Munari et al. (2002) analyze the impact of privatization on corporate R&D and companies' innovative behavior and found that privatization is followed by a reduction in R&D spending and a shift towards more commercially-oriented projects. Innovation activities benefits the society. So, state owned enterprises have the incentives over non- state-owned enterprises to further social welfare by investing into more innovation projects. Also, analysts' pressure on privatized firms to meet the short-term expectations might thwart some innovation projects that might be beneficial to the firm in the long run, but it is detrimental in the short run.

The article by Boubakri et al. (2013) does not specifically relate privatization to innovation but rather risk-taking behaviour. This article helps to gain insight on innovation indirectly. Innovation is usually preceded by years of expensive R&D and their high costs to create is an additional source of risk as investments must be justified with positive expected future returns. The authors suggest that privatization is linked to improvements in corporate governance, openness in foreign investments and that these are key determinants in corporate risk-taking.

Hypotheses

The ownership structure changes because of the deliberate sale of assets to private economic agents. Privatization to foreign owners leads to more restructuring (Djankov and Murrell (2002) and Estrin et al. (2009)). Such restructuring by more risk-oriented investors who are more likely to adopt innovative practices and hence venture into more risky projects leads to increased earnings volatility after the divestiture. Boubakri et al. (2013) postulate that public enterprises are inefficient because the objective of the state-owned firm is not to maximize the profits or shareholder value but rather maximizing employment and regional development. As a result, state owners are less likely to maximize profits by cost cutting measures or venturing into risky projects that may induce an unfavorable swing in voters' opinions on the current government. Therefore, government will influence firms to be conservative to have maximum employment and hence maximum support from the voters. Another issue related to agency problem is the lack of adequate monitoring in state owned enterprises. As there is no individual owner with incentives to engage in active monitoring, the managers are not actively monitored. As a result, managers divert corporate resources for private benefits and prevent the firm from undertaking risky projects. After privatization, these issues are mitigated since the managers can now engage freely into risky ventures. Innovation is a high-risk endeavor with a low likelihood of being turned into an economically viable product. To gain a competitive advantage in the market, companies that are more risk-oriented would boost their capital allocation to innovative working techniques, goods, services, and processes. So, in a post-privatization dataset, we hypothesize the following:

Hypothesis 1: Post-privatized companies that are more risk-taking should be expected to engage in more innovation activities.

The degree to which a firm collaborates with other firms affects its innovation activities. It is very likely that private firms are more open and willing to collaborate with other organizations. As private firms are more open to collaborate with other external partners, they increase their innovative performance. State owned enterprises, due to political reasons, are likely to conduct their research and development activities in house rather than collaborating with external organizations (Somé et al. (2021)). Due to the above reasons, the number of patents associated with post-privatization firms are more. Also, due to the increase in cross-country collaboration in innovation activities, the firms' performance is improved. The increase in the number of patents and the dispersion of knowledge sourcing enhances the firms' returns of assets, and greater sales growth but only after privatization (Somé et al. (2021)). Since, collaboration activities help companies to remain competitive in today's environment, by collaborating, a firm can increase its innovation capabilities. Hence, we hypothesize the following:

Hypothesis 2: More collaboration would lead to more innovation activities

As shown in Boubakri et al. (2013), public enterprises are risk averse as the objective of the state-owned firm is not to maximize profits or shareholder value but rather maximizing employment and regional development. Privatized companies don't have such objectives and tend to venture into more risky projects. Also, Boubakri et al. (2013) posits that state-owned firms are inefficient because their managers are not adequately monitored due to the lack of individual owner. Hence, post-privatization firm-value enhancement, which can be reached by undertaking risky projects, may not be achieved. Hence, we hypothesize the following:

Hypothesis 3: State ownership is negatively related to the innovation activities of the firm.

Measures of corporate innovation activities

Following the innovation literature (e.g., Cho et al. (2016); Chang et al. (2015)), we can use patent-based metrics as a dependent variable that will serve as a proxy for innovation. The two types of metrics that are commonly used are citations-counts and patents. Since patents vary in their economic and technological significance, measuring innovation activity by the number of patent counts does not reveal much information about the companies' innovation activities (Chkir et al. (2021)). Another benefit of using patents' citations-count is that the patent can still receive citations even after the firm goes bankrupt, which makes it less likely to suffer from survivorship bias. Due to the above-mentioned reasons, citations-count serves as the best proxy for measuring the companies' innovation activities and hence the primary measure of innovation. Companies' patent counts can serve as a secondary measure and can be used to check for robustness once we have performed the basic analysis using citations count.

Measures of corporate risk taking

Following previous studies, we will use the risk measures used in the paper by Boubakri et al. (2013). The primary measure of corporate risk-taking is the volatility of a firm's earnings over four years of overlapping periods. If we take a period of 7 years after privatization, then volatility is computed as volatility of earnings from +1, +4; +2, +5; +3, +6; +4, +7.

Here the earnings are defined as earnings before interest and taxes (EBIT) over total assets. For robustness, we will also use other risk measures defined in the paper Boubakri et al. (2013). The primary measure is named as Risk1 and other risk measures are defined in section 3.2.

Firm Level Control Variables

To separate the effects of the incremental explanatory power of the firms' risk-taking behaviour on the corporate innovation, it is important to identify some control variables.

- Size

  According to Symeonidis (1996), innovation increases with firm size because of the following reasons: -

  Due to the fixed nature of the research and development costs, it can be covered only if the sales are large enough. These costs are mostly independent of the size of the market. Since innovation is a risky activity with very uncertain outcomes, it is unlikely that small firms can engage into extensive innovation activities. If the outcomes of the research and development activities are not as expected, it might hamper the long-term prospects of a small firm, given the fixed nature of the research and development costs.

  Also, the cost of research and development activities depends on the industry as well. If we look at the airlines industry, then some large firms also fail to carry out extensive innovation activities because of the level of research costs in the industry. Also, it might be considered that small firms might collaborate into a research activity which might compete with a large firm. But sometimes, management of a group of firms might be difficult and might not lead to a fruitful result.

  We also observe the increase in research and development costs across industries. The increase in the costs is not uniform over industries. So, if the increase in research costs increases at a rate greater than the sales growth in the industry, small firms can be less prone to innovate.

  Research and development costs are sunk costs, meaning the costs are incurred regardless of whether the firm makes a profit out of it. Due to the uncertainty of the outcome, R&D activities carry huge amount of risk. So, the availability of finance will decide the extent of innovation activities the firm can undertake. Specifically, the availability of external or internal finance will determine the firms' engagement in innovation activities. Because of this, large firms tend to innovate more because of better access to external finance.Also, there are scale and scope economies in the production of innovation. Also, large firms are better poised to exploit innovation opportunities. Large firms can take many projects at one time and hence spread the risks of R&D.

- Return on Assets

  Return on Assets (ROA) is defined as the EBIT/ total assets. It can be used as a proxy for profitability of a company. All else equal, the more profitable the company is, the more it will have money to spend on innovative activities. So, we expect ROA/profitability of a company to be positively related to its innovation activities.

- Plant property and equipment

  Plant property and equipment is defined as the plant property and equipment divided by the total assets of the company. All else equal, the more the company invests in fixed assets, the less it will have money to spend on innovative activities. So, we expect plant property and equipment of a company to be negatively related to its innovation activities of the firm.

- Leverage

  Leverage is defined as the total debt divided by the total assets of the company. All else equal, the more debt the company has, the less money to spend on innovative activities because of cash outflows due to debt servicing (interest and principal repayment). So, we expect leverage to be negatively related to the innovation activities of the firm.

- Capital Expenditures

  Capital expenditures is defined as the ratio of capital expenditures to total book value of assets.All else equal, the more capital expenditures the company has, the less money to spend on innovative activities. So, we expect capital expenditures to be negatively related to the innovation activities of the firm.

- Cash

Cash is defined as the cash divided by the total assets of the company. Adler et al. (2019) argues that cash and cash equivalents impact the innovation activities of the firms. The investment in innovation is subject to liquidity shocks before any fruitful outcome, which one might expect from successful innovation. So, the more cash the company has, the more it is going to invest in innovative activities.

- Dividend

  Dividend is defined as the dividend paid out by the company divided by the total assets of the company. All else equal, the more dividend the company pays out to its shareholders, the less cash it has to invest into innovative activities. So, we expect a negative relationship between dividend and innovation.

Section 1

Data Collection

The final database is formed by merging three different databases from three different sources. The three databases are as follows: 1) Database with ownership structure of the companies 2) Database with citation activity of the firm 3) Database with firm- specific control variables.

1. **Database with ownership structure of the companies**

   This database is provided by Narjess Boubakri and has been used in Boubakri et al. (2013). Since, the relationship between risk taking and innovation activities is done in a post-privatized companies' data, this database provides the list of companies that are privatized. A brief description of the relevant variable is as follows:

   - stateshares1stclass

     The number of first class shares owned by state owned enterprise (SOE).

   - state1stclass

     The percentage of first class shares owned by state owned enterprise (SOE).

2. **Database with citation activity of the firm** The citations that the company's patents received are used as a proxy for the firm's innovation activities. Using the list of the privatized firms obtained from the ownership database, companies' citations are

found from Derwent Innovation. A brief description of some of relevant variables obtained from the Derwent Innovation database is as follows:

- Title

  A concise descriptive English-language title written by Derwent Abstractors to highlight the content and novelty of the invention disclosed in the patent specification.

- Patent Number

  The listed patent number(s) are for all members of the patent family.

- Patent Assignee Name(s) and Code(s)

  The individual(s) or corporate body to whom all or limited rights of the patent are legally transferred, along with a unique four-letter code assigned by Derwent.

- Citing Patents

  Displays the number of patent family records whose members have cited members of the current patent family. A zero means that no patents covered in the current database cite members of this patent family. This variable is used as a proxy for innovative activities of the firm and is the dependent variable in the regression.

- Patents Cited by Inventor / Examiner

  Displays the number of patents cited by the inventor / examiner. A zero means there are no patent references, or the references were not keyed into the database.

- Articles Cited by Inventor / Examiner

  Displays the number of articles (non-patent items) cited by the inventor / examiner. A zero means the patent has no article references or the references were not keyed into the database.

- Optimized Assignee and Patent Assignee name

Identifying the true owner of the patent is important to calculate the citations received on those patents. But sometimes, it can be unclear who owns the patent. The name on the document is just the beginning as it may or may not be accurate. The inaccurate name on the patent document may arise due to typographical errors, a variant of the common business name is written, or it may be a subsidiary company or in worse case it may be absent altogether on the patent document. For example: IBM corporation can be referred as Int Business Mashines (a typographical error), or International Business Machines, IBM Res Gmbh (common business variant names), or Netezza corp. (a subsidiary of IBM).

To address this issue, Derwent innovation has created a separate field called "optimized assignee". Optimized Assignee provides a single preferred entity name which is not only a normalized company name, but also predicts missing assignees, and takes corporate structure into account. It includes the probable Assignee (where no organization is listed on the application) and considers the latest reassignment, company hierarchy, and name clean-up/normalization.

So, the citations data of each firm is searched using the optimized assignee field of Derwent Innovation. The data obtained is of various patents filed by the company in a particular year. The citations received on each the patents are given by the field "Count of Citing Patents". To get the total number of citations received by a particular company each year, the citations received on each of the patents are summed up over the year. For example, if Adobe Inc. filed for two patents in the year 2007, the citations received on the first patent is 10 and the citations received on the second patent is 20 then the total number of citations received by Adobe Inc. in the year 2007 is 30. This process is repeated for each of the companies and for each year to find the total number of citations received by a company over a period of several years. This can be taken as the proxy for the innovation activity of the firm, and this is the dependent variable in our equation.

- Average of inventor count

  This is our primary measure for collaboration. The data on the number of inventors that worked on a particular patent is taken from the Derwent Innovation database. Since in a given year, a company files for multiple patents with varying number of inventors working on every patent, we took the average number of inventors over one year. For e.g., if Adobe Inc. files for 2 patents in the year 2007 and the number of inventors that worked on the first patent is 2 and the number of inventors that worked on the second patent is 4, then the average number of inventors working on a patent is taken as 3 for the year 2007 for Adobe Inc.

- Average of assignee count

  This is one additional measure of collaboration used for robustness check. The total number of companies that filed for the patent is defined as the assignee count. The data on the number of assignees is taken from the Derwent Innovation database. Similar to inventor count, a patent can have multiple assignees. Since, there are multiple patents published in a particular year, we took the average number of assignees over one year. As discussed above, the patents were searched using the optimized assignee field of the Derwent Innovation database. This yields all the patents filed by a company in any given year. This also includes the patents in which the company has collaborated with other companies and filed jointly for the patent rights. So, every patent has a number of assignee count which corresponds to the number of entities which were granted the patent rights. For e.g., if Adobe Inc. files for 2 patents in the year 2007 and the number of assignees that filed for the first patent is 2 and the number of assignees that filed for the second patent is 1, then the average number of assignees that filed for a patent is taken as 1.5 for the year 2007 for Adobe Inc.

3. **Database with firm- specific control variables**

   The independent variable in the equation is the volatility in return. The return is defined as the Earnings before interest and taxes (EBIT)/ total assets. Along with these, there are some other firm level control variables that were collected. The data firm level control variables were collected from the compustat database. A brief description of the control variables is as follows:

   - at - total assets
   - xrd - research and development
   - capx - capital expenditure
   - ch - cash
   - dlc - Debt in current liabilities
   - dltt - total long-term debt
   - dvc - dividends common
   - ebit - earnings before interest and taxes
   - ppent - total property plant and equipment

Sample Description

To investigate the effects of risk taking, collaboration and state ownership on innovation, we construct a final sample by merging all the three databases mentioned above. To be included in the sample, we require each firm's data be available for at least seven consecutive years in order to compute the volatility measure.

The final sample consists of 246 privatized firms from 19 countries. The detailed description is shown in Table 1.2. Table 1.1 presents the descriptive statistics by industry for the 246 privatized firms considered in the study. The sample thus has variation across industries. The final data is from 2007 to 2015.

Table 1.1 – Description of the sample of newly privatized firms by industry

Industry	Number	Percentage
Basic Materials	22	8.94
Consumer Goods	39	15.85
Consumer Services	4	1.63
Health Care	37	15.04
Industrials	48	19.51
Oil & Gas	21	8.54
Technology	65	26.42
Telecommunications	9	3.66
Utilities	1	0.41
Total	246	100

Table 1.2 – Description of the sample of newly privatized firms by country

Country	Number	Percentage
Australia	2	0.81
Belgium	1	0.41
Brazil	1	0.41
Canada	1	0.41
Denmark	1	0.41
Finland	4	1.63
France	12	4.88
Germany	8	3.25
Italy	4	1.63
Japan	16	6.5
Netherlands	4	1.63
Norway	3	1.22
Russia	1	0.41
South Korea	3	1.22
Spain	2	0.81
Sweden	6	2.44
Switzerland	6	2.44
United Kingdom	13	5.28
United States	158	64.23
Total	246	100

Section 2

Methodology

Modelling count data

Count is typically used to enumerate units, things, or events. We might have count to describe the number of bicycles rides, number of rides taken by the passengers etc. A count variable only takes non-negative integer values. Also, the values are assumed to be independent from each other.

All parametric statistical models are based on an underlying probability distribution. In a simple linear or multiple linear regression model, the assumption is that the error term is normally distributed. When we are estimating a least square model, we are trying to estimate parameters of the underlying probability distribution that best characterizes the data. Similarly, when we are trying to model the count data, we are estimating the parameters of a probability distribution that best characterizes or represents the data. Also, the data being modelled is assumed to be a random sample of a greater population of data. The probability distribution that we are modelling is assumed to describe the population data and not the sample data.

Generally, count data models are based on two probability distributions. The Poisson and the negative binomial probability distribution functions. Since the dependent variable in our case is the count of citing patents, exploring these two models in detail becomes necessary.

2.1 Poisson regression model

Poisson regression models are the most basic forms of count models. The key underlying assumption is that the count data is structured in the form of a Poisson probability distribution function. In a Poisson distribution, only a single parameter, mean, is estimated, since the Poisson distribution assumes that the mean and the variance are the same. Apart from the equi- dispersion (mean = variance), Poisson also makes a few other assumptions as discussed in Hilbe (2014):

- The distribution is discrete with a single parameter, the mean, which is usually symbolized as either lambda or mu. The mean is also understood as the rate parameter. It is the expected number of times that an item or event occurs per unit of time, area, or volume.
- The response terms, or y values, are non-negative integers, i.e., the distribution allows for the possibility of counts where y>=0.
- Observations are independent of one another.
- No cell of observed counts has substantially more or less than what is expected based on the mean of the empirical distribution. For example, the data should not have more zero counts than is expected based on a Poisson distribution with a given mean. As the value of mu increases, the probability of zero counts is reduced.
- The mean and the variance of the model are identical, or at least nearly the same, i.e., Poisson distributions with higher mean values have correspondingly greater variability.
- The Pearson Chi2 dispersion statistic has a value approximating 1.0.

The equi-dispersion criteria, i.e., the mean = variance in the data is seldom satisfied. There is usually some level of under or over dispersion in the data (Hilbe (2014)). In an under dispersed data, the mean is greater than the variance in the data. In an over dispersed data, the variance is greater than the mean. In most of the cases, we have the problem of over dispersion. In the case of an over dispersed model, the standard errors tend to be biased and hence, unreliable. In order to correctly model an over dispersed data other models such as negative binomial model can be implemented.

The Poisson distribution has the following Probability Mass Function (Cameron and Trivedi (2013)).

$$P(k) = \frac{e^{-(\lambda * t)} * (\lambda * t)^k}{k!} = Poisson(\lambda * t) \tag{2.1}$$

P (k) - Probability of seeing k events in time t

k - events

λ - Event rate (number of events per unit time)

t - time

The expected value or the mean for a Poisson Distribution is λ. Thus, in absence of any information, we can expect to see the mean number of events or λ events in each time t. The time interval can be 1 hour, 1 day etc. So, for any interval t, we can expect to see λt events.

A Poisson regression model for a non-constant λ:

A common situation is to model a count data where λ can change from one observation to next. Here we assume that the value of λ is influenced by a vector of explanatory variables, also known as regression variables X.

In a Poisson regression model, the event counts y is assumed to be Poisson distributed, which means the probability of observing y is a function of the event rate vector λ. The job of the Poisson Regression model is to fit the observed counts y to the regression matrix X via a link-function that expresses the rate vector λ as a function of 1) the regression coefficients β and 2) the regression matrix X.

The link function that connects the λ with the regression matrix X and the regression coefficients β is the exponential function.

$$\lambda = e^{X\beta} \tag{2.2}$$

λ – event rates (a n*1 matrix)

X - regression matrix (a n*m matrix)

β – regression coefficients (a m*1 matrix)

n – total number of observations

m – total number of independent variables

The link function works best because it keeps λ non-negative even when the regressors X or the regression coefficients β have negative values. This satisfies the requirement of the count-based data that it should only take non-negative values.

The complete specification of the Poisson regression model is given as:
For the i^{th} observation in the data denoted by y_i corresponding to the row of regression variables x_i, the probability of observing count y_i is given as:

$$P(y_i|x_i) = \frac{e^{-\lambda_i} * \lambda_i^{y_i}}{y_i!} \tag{2.3}$$

P($y_i|x_i$) – Probability of seeing count y_i given the regression vector x_i.
λ_i – event rate for the i^{th} observation.

Where λ_i is given as follows:

$$\lambda_i = e^{x_i * \beta} \tag{2.4}$$

λ_i - event rate for i^{th} observation.

x_i - i^{th} row of the regression matrix X (a 1*m matrix).

β_i - regression coefficients (a m*1 matrix).

Identifying the regression coefficients β

The values of the regression coefficients are computed such that it would make the vector of observed counts y most likely. The technique for identifying the coefficients β is called the Maximum Likelihood Estimation.

Our assumption is that the count data arise from a Poisson process. Hence, we can say that their probabilities of occurrence are given by the Poisson probability mass function (PMF).

$$PMF(y_i|x_i) = \frac{e^{-\lambda_i} * \lambda_i^{y_i}}{y_i!} \tag{2.5}$$

The probabilities for the first two cases:

$$PMF(y_1|x_1) = \frac{e^{-\lambda_1} * \lambda_1^{y_1}}{y_1!} \tag{2.6}$$

$$PMF(y_2|x_2) = \frac{e^{-\lambda_2} * \lambda_2^{y_2}}{y_2!} \tag{2.7}$$

The lambda (λ_1, λ_2) is given as:

$$\lambda_1 = e^{x_1 * \beta} \tag{2.8}$$

$$\lambda_2 = e^{x_2 * \beta} \tag{2.9}$$

Here, x_1, x_2 are the first two rows of the regression matrix. Since the counts y are Poisson distributed, $y_1, y_2, y_3, \ldots, y_n$ are independent random variables. Hence the joint probability of occurrence of $y_1, y_2, y_3, \ldots, y_n$ can be expressed as a simple multiplication of the individual probabilities.

$$P(y|X) = P(y_1|x_1) * P(y_2|x_2) * P(y_3|x_3) * P(y_4|x_4) * \ldots\ldots\ldots\ldots\ldots\ldots * P(y_n|x_n) \tag{2.10}$$

The joint probability after plugging in the individual count probabilities.

$$P(y|X) = \frac{e^{-\lambda_1} * \lambda_1^{y_1}}{y_1!} * \frac{e^{-\lambda_2} * \lambda_2^{y_2}}{y_2!} * \ldots\ldots * \frac{e^{-\lambda_n} * \lambda_n^{y_n}}{y_n!} \tag{2.11}$$

The likelihood function for β is given by:

$$L(\beta) = P(y|X) = \frac{e^{-\lambda_1} * \lambda_1^{y_1}}{y_1!} * \frac{e^{-\lambda_2} * \lambda_2^{y_2}}{y_2!} * ...... * \frac{e^{-\lambda_n} * \lambda_n^{y_n}}{y_n!} \tag{2.12}$$

What value of β will make the given set of observed counts y most likely? It is the value of β for which the joint probability shown in the above equation achieves the maximum value. Since we want to find the maximum value, we can find it by differentiating the equation w.r.t β and setting the differential equation to 0. It is easier to differentiate the logarithm of the joint probability function. The solution to the logged equation yields the same optimal value of β. This logarithmic equation is called the log-likelihood function and it is given as follows:

$$\ln L(\beta) = \sum_{i=1}^{n} (y_i * x_i * \beta - e^{x_i * \beta} - \ln y_i!) \tag{2.13}$$

$$\frac{\partial \ln L(\beta)}{\partial \beta} = \sum_{i=1}^{n} (y_i - e^{x_i * \beta}) * x_i = 0 \tag{2.14}$$

Solving this equation for the regression coefficients β will yield the Maximum Likelihood Estimate (MLE) for beta. This can be done using the Iteratively Reweighted Least Squares (IRLS), which is available in the python package statsmodels.

2.2 Negative Binomial Regression Model

A negative binomial regression is an extension of the Poisson regression model. It relaxes the assumption of mean equals to variance of the Poisson regression model. It requires us to define a new parameter α which is used to express the variance in the terms of means as follows:

$$Variance = mean + \alpha * (mean)^p \tag{2.15}$$

When p = 1,

$$
\begin{aligned}
Variance &= mean + \alpha * mean \\
&= (1+\alpha) * mean
\end{aligned} \tag{2.16}
$$

This case is referred to as the NB1 model or a linear negative binomial model.

When p = 2,

$$Variance = mean + \alpha * (mean)^2 \tag{2.17}$$

This case is referred to as the NB2 model or a quadratic negative binomial model. Since the variance is much greater than mean in our data, a NB2 model is a better choice over a NB1 model.

The traditional binomial model or a NB2 model has the same distributional assumptions as a Poisson distribution with some exceptions. As we can see in Equation 2.17, two parameters affect the variance of the distribution. So, apart from the mean, the variance is also affected by the dispersion parameter(α) multiplied by the square of the mean. Because of this, the negative binomial model allows us to model a far wider range of variability than the Poisson model. Because of the greater value of the variance than the mean in a negative binomial model, it is nearly always used to estimate the parameters of over dispersed data. Negative binomial model (NB2) makes several assumptions as discussed in Hilbe (2014):

- The dependent variable, y, is a count consisting of nonnegative integers.
- As the value of λ increases, the probability of 0 counts decreases.
- y must allow for the possibility of 0 counts.
- The fitted variable, λ, is the expected mean of the distribution of y.
- The variance is closely approximated as mean+ $\alpha * (mean)^2$.
- A foremost goal of NB regression is to model data in which the value of the variance exceeds the mean, or the observed variance exceeds the expected variance.

- The model is not misspecified.

A negative binomial model is a two-parameter model. The first parameter is the mean and the second parameter being the dispersion parameter(α).

Finding the value of α for a NB2 model.

When a negative binomial is estimated using a full maximum likelihood algorithm, both mean and the dispersion parameter α are estimated. When it is estimated using a generalized linear model algorithm, only mean is estimated; α must be inserted into the algorithm as a constant.

We are implementing our negative binomial regression using Python's Generalized Linear Models module. Hence, the dispersion parameter α needs to be inserted into the algorithm as a constant.

One method to compute α is suggested by Cameron and Trivedi (2013). For a NB2 model, they suggest a means to calculate α using an auxiliary OLS regression without a constant.

$$\frac{(y_i - \lambda_i)^2 - \lambda_i}{\lambda_i} = \alpha * \lambda_i + 0 \tag{2.18}$$

Variables in the regression:
y_i= dependent count variable
λ_i = the vector of event rates

$$Y = B_1 * x + B_0 \tag{2.19}$$

$Y = \frac{((y_i - \lambda_i)^2 - \lambda_i)}{\lambda_i}$
$B_1 = \alpha$
$x = \lambda_i$
B_0 = intercept of the regression

The value of α for a NB1 model can be found by performing the following regression.

$$\frac{((y_i - \lambda_i)^2 - \lambda_i)}{\lambda_i} = \alpha + 0 \tag{2.20}$$

Section 3

Risk Taking and Innovation

In this chapter, we present our analysis of the effects of corporate risk-taking on innovation. We are using two different models to analyze the effects. First, we perform a Poisson regression model. Acknowledging the over dispersion in our data, we then implement a negative binomial model.

3.1 Risk measure 1

Our primary risk measure come from Boubakri et al. (2013).The risk measure, Risk1, is defined as the volatility of return on assets over a rolling period of 4 years. Here the return on assets are defined as the ratio of earnings before interest and taxes (EBIT) and total assets (At).

$$Risk1 = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}(E_{i,t} - \frac{1}{T}\sum_{t=1}^{T}E_{i,t})^2 | T = 4} \tag{3.1}$$

Where $E_{i,t} = \frac{EBIT_{i,t}}{A_{i,t}}$

$EBIT_{i,t}$ is defined as the earnings before interest and taxes for firm i in year t. $A_{i,t}$ is defined as the total assets for firm i in year t; T over (+1 to +4, +2 to +5, +3 to +6, ...)

Table 3.1 – Descriptive Statistics

This table provides summary statistics for all the variables used in the main regression analysis of the impact of risk taking on corporate innovation.

Variables	**Count**	**Mean**	**Variance**
Citations Count	1375	10088.86	4.91e+08
Risk1	1375	0.027	7.16e-04
size	1375	9.620	2.10e+00
Return on Assets	1375	0.153	6.66e-03
Plant Property & Equipment	1375	0.213	2.03e-02
Leverage	1375	0.224	2.45e-02
Capital Expenditures	1375	0.043	1.06e-03
Cash	1375	0.117	9.41e-03
Dividend	1375	0.023	9.34e-04

3.1.1 Regression Model Implementation

Table 3.1 provides the summary statistics of the dependent, independent, and control variables. There are about 10088 citations per firm per year, on average. The citations data displays a significant variability as evidenced in a high variance of 49100000. Our independent variable, Risk1, has a mean of 0.0267 and displays a very low variability. On average, as a proportion of total assets, profitability account for 15%, property,plant and equipment for 21%, leverage for 22%, capital expenditures for 4%, cash for 11%, and dividends for 2%. The summary statistics of control variables is approximately similar to Chkir et al. (2021).

This section investigates the relation between risk taking and firm innovation by using the Poisson regression model. Table 3.2 portrays the results from the regression analysis of the relation between risk taking and corporate innovation.

Table 3.2 – The impact of Risk1 on innovation using Poisson Model

This table represents the results of the Poisson regression for risk measure 1. The dependent variable is the citation counts received by the firm on its published patents. The risk measure, risk1, corresponds to the volatility of ROA(return on Assets) computed over a rolling period of 4 years. Here ROA is defined as the EBIT(Earnings before Interest and Taxes)/ At(total assets). The control variables are size (defined as log of total assets), return on assets (defined as Earnings before interest and taxes/total assets), Plant Property and Equipment (defined as PPE/total assets), Leverage (defined as total debt/total assets), Capital Expenditures (defined as capital expenditures/total assets), Cash (defined as cash/total assets), Dividend (defined as dividend/total assets). Our sample consists of the data on private firms. The citations data is obtained from the Derwent Innovation Database. The control variables are obtained from the compustat database. The data is from 2007 to 2015. To compute the standard deviation over a rolling period of 4 years only the companies that have more than 7 years of data are kept. The citations data and the data on control variables are merged in the final table resulting into 1375 firm-year observations.

Df Residuals:	1366	**Pearson chi2:**	2.93e+07			
	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	2.4222	0.003	872.698	0.000	2.417	2.428
Risk1	0.6315	0.011	55.071	0.000	0.609	0.654
size	0.7358	0.000	3041.175	0.000	0.735	0.736
Return on Assets	1.3834	0.004	312.333	0.000	1.375	1.392
Plant Property & Equipment	-6.6424	0.004	-1857.047	0.000	-6.649	-6.635
Leverage	-2.7192	0.002	-1197.438	0.000	-2.724	-2.715
Capital Expenditures	17.6001	0.013	1381.035	0.000	17.575	17.625
Cash	1.5869	0.004	451.458	0.000	1.580	1.594
Dividend	-1.9357	0.012	-162.372	0.000	-1.959	-1.912

As expected, since the dependent variable's variance is much greater than the mean, the model is over-dispersed. Consequently, we can see that the standard errors are underestimated resulting into inflated z-values which we can see in Table 3.2. We also performed some additional tests to check whether the model is over dispersed or not. One test commonly used to check for over-dispersion is the Pearson Chi2 test.

Pearson Chi2 Statistic

As defined in Hilbe (2014), the Pearson Chi2 statistic is the sum of squared residuals weighted or adjusted by the model variance. As seen in Table 3.2, the value of the Pearson Chi2 for the Poisson model is 29300000.

The dispersion statistic is defined as the ratio of Pearson Chi2 statistic and the residual degrees of freedom.

$$dispersion\ statistic = \frac{Pearson\ Chi2\ statistic}{residual\ degrees\ of\ freedom} \tag{3.2}$$

The dispersion statistic should have a value of 1. Values greater than 1 indicate an over dispersed model. Similarly, values less than 1 indicate an under dispersed model.

Taking the data from Table 3.2, the dispersion statistic for our model is:

$dispersion\ statistic = \frac{Pearson\ Chi2\ statistic}{residual\ degrees\ of\ freedom} = \frac{29300000}{1366} = 21449.48$

The dispersion statistic of 21449.48 indicates a highly over dispersed model. To overcome the problem of over dispersion in the data, negative binomial regression is implemented.

Negative Binomial Model

As discussed in section 3.1.1, the dispersion statistic is much higher than 1, our data is highly over dispersed. So, the negative binomial 2 model is an appropriate choice as the relationship between mean and variance is depicted using a quadratic function in a negative binomial 2 model.

The negative binomial regression results for different values of α is shown in Table 3.3.

Table 3.3 – Negative Binomial Regression for different values of alpha

This table represents the results of the Negative Binomial regression for risk measure 1 at different values of α. The dependent variable is the citation counts received by the firm on its published patents.The risk measure, risk1, corresponds to the volatility of ROA(return on Assets) computed over a rolling period of 4 years.Here ROA is defined as the EBIT/At(total assets). The control variables are size (log of total assets), Plant Property and Equipment (PPE/total assets), Leverage (total debt/total assets), Capital Expenditures (capital expenditures/total assets), Cash (cash/total assets), Dividend (dividend/total assets). Our sample consists of the data on private firms. The citations data is obtained from the Derwent Innovation Database. The control variables are obtained from the compustat database. The data is from 2007 to 2015. To compute the standard deviation over a rolling period of 4 years only the companies that have more than 7 years of data are kept. The citations data and the data on control variables are merged in the final table resulting into 1375 firm-year observations. *** , ** , and * denote significance at the 1%, 5%, and 10% levels, respectively. † - α value obtained from the auxiliary OLS regression. Standard errors in parenthesis.

	$\alpha = 1.212^{\dagger}$	$\alpha = 2.0$	$\alpha = 2.5$
const	2.77***	2.77***	2.77***
	(0.25)	(0.32)	(0.36)
Risk1	6.67***	6.67***	6.67***
	(1.22)	(1.57)	(1.76)
Capital Expenditures	10.83***	10.83***	10.83***
	(1.34)	(1.72)	(1.92)
Cash	1.41***	1.41***	1.41***
	(0.36)	(0.47)	(0.52)
Dividend	-2.07*	-2.07	-2.07
	(1.07)	(1.37)	(1.53)
Leverage	-1.93***	-1.93***	-1.92***
	(0.20)	(0.26)	(0.29)
Plant Property & Equipment	-3.84***	-3.84***	-3.84***
	(0.32)	(0.41)	(0.45)
Return on Assets	-0.02	-0.02	-0.02
	(0.40)	(0.52)	(0.58)
size	0.67***	0.67***	0.67***
	(0.02)	(0.03)	(0.03)
N	1375	1375	1375
Pearson chi2:	2.83e+03	1.72e+03	1.37e+03
Df Residuals:	1366	1366	1366

- The dispersion statistic for the all the three negative binomial models using the

Pearson Chi2 and residual degrees of freedom obtained from Table 3.4 is as follows:

1. $\alpha = 1.212$

   $$dispersion\ statistic = \frac{Pearson\ Chi2\ statistic}{residual\ degrees\ of\ freedom} = \frac{2830}{1366} = 2.07$$

2. $\alpha = 2.0$

   $$dispersion\ statistic = \frac{Pearson\ Chi2\ statistic}{residual\ degrees\ of\ freedom} = \frac{1720}{1366} = 1.259$$

3. $\alpha = 2.5$

   $$dispersion\ statistic = \frac{Pearson\ Chi2\ statistic}{residual\ degrees\ of\ freedom} = \frac{1370}{1366} = 1.0029$$

- we can see that the dispersion statistic is still much greater than 1 for α of 1.21 and 2.0 . The model is properly specified at the α value of 2.5 as dispersion statistic is closest to 1 in this case. The detailed summary for negative binomial regression model computed using α of 2.5 is shown in Table 3.4.

- The dispersion statistic is also much less than the dispersion statistic obtained from the Poisson model. The dispersion statistic of the Poisson model is 21449.48 which is much greater than 1.0029 obtained from the negative binomial model with $\alpha =$ 2.5.

- As we can see in Table 3.4, the standard errors are much greater than the standard errors of the Poisson Model shown in Table 3.2. Consequently, the z-values of the negative binomial model are much less than the z-values obtained in the Poisson Model (shown in Table 3.2).

Table 3.4 – The impact of Risk1 on innovation using NB2 Model

This table represents the results of the Negative Binomial regression for risk measure 1 at $\alpha = 2.5$. The dependent variable is the citation counts received by the firm on its published patents.The risk measure, risk1, corresponds to the volatility of ROA(return on Assets) computed over a rolling period of 4 years. Here ROA is defined as the EBIT/ At(total assets). The control variables are size (log of total assets), Plant Property and Equipment (PPE/total assets), Leverage (total debt/total assets), Capital Expenditures (capital expenditures/total assets), Cash (cash/total assets), Dividend (dividend/total assets). Our sample consists of the data on private firms. The citations data is obtained from the Derwent Innovation Database. The control variables are obtained from the compustat database. The data is from 2007 to 2015. To compute the standard deviation over a rolling period of 4 years only the companies that have more than 7 years of data are kept. The citations data and the data on control variables are merged in the final table resulting into 1375 firm-year observations.

	Df Residuals:	1366	**Pearson chi2:**	1.37e+03		
	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	2.7721	0.362	7.647	0.000	2.062	3.483
Risk1	6.6715	1.757	3.798	0.000	3.228	10.115
size	0.6685	0.034	19.898	0.000	0.603	0.734
Return on Assets	-0.0234	0.577	-0.041	0.968	-1.154	1.107
Plant Property & Equipment	-3.8423	0.453	-8.476	0.000	-4.731	-2.954
Leverage	-1.9250	0.286	-6.725	0.000	-2.486	-1.364
Capital Expenditures	10.8340	1.924	5.631	0.000	7.063	14.605
Cash	1.4125	0.521	2.712	0.007	0.392	2.433
Dividend	-2.0657	1.534	-1.347	0.178	-5.072	0.941

Table 3.5 – The impact of Risk1 on innovation using OLS

This table represents the results of the OLS regression for risk measure 1. The dependent variable is the citation counts received by the firm on its published patents. The risk measure, risk1, corresponds to the volatility of ROA(return onAssets) computed over a rolling period of 4 years. Here ROA is defined as the EBIT/At(total assets). The control variables are size (defined as log of total assets), Plant Property and Equipment (defined as PPE/total assets), Leverage (defined as total debt/total assets), Capital Expenditures (defined as capital expenditures/total assets), Cash (defined as cash/total assets), Dividend (defined as dividend/total assets). Our sample consists of the data on private firms. The citations data is obtained from the Derwent Innovation Database. The control variables are obtained from the compustat database. The data is from 2007 to 2015. To compute the standard deviation over a rolling period of 4 years only the companies that have more than 7 years of data are kept. The citations data and the data on control variables are merged in the final table resulting into 1375 firm-year observation.

Df Residuals:	1366	**Df Model:**	8	
	coef	**std err**	**z**	**P> |z|**
const	-4.535e+04	4680.648	-9.690	0.000
Risk1	3.502e+04	1.59e+04	2.200	0.028
size	6341.1252	523.977	12.102	0.000
Return on Assets	1.268e+04	7871.048	1.611	0.107
Plant Property & Equipment	-4.992e+04	5425.311	-9.202	0.000
Leverage	-1.872e+04	4005.739	-4.672	0.000
Capital Expenditures	1.276e+05	1.94e+04	6.575	0.000
Cash	1.084e+04	4807.205	2.254	0.024
Dividend	-1.809e+04	1.42e+04	-1.270	0.204

Results in Table 3.4 show the evidence of the effects of risk taking on firms' innovative activities. The results suggest that risk taking is positively associated with innovation activities and lend empirical support to the view that risk taking fosters corporate innovation. This association is statistically significant at 1% level. We also performed an OLS regression analysis as shown in Table 3.5. In line with the expectations, these results support hypothesis 1, which posits that more risk should lead to more innovation. As shown in Boubakri et al. (2013), public enterprises are risk averse as the objective of the state-owned firm is not to maximize profits or shareholder value but rather maximizing employment and regional development. Privatized companies don't have such objectives and tend to venture into more risky projects. Innovation is a risky endeavor with very slim chances that it would be converted into a commercially viable product. Companies that are more risk-oriented would increase their capital allocation to new working methods, products, services and processes in order to have a competitive edge in the market.

We can also see that the coefficient of risk 1 is 6.67 which implies that risk taking has a significant impact on citations count of the firm. Since, the link function between the independent variable and the dependent variable in a negative binomial is a log function, we can say that 1 unit increase in risk will increase the log of citations count by 6.67, holding other variables constant. In other words, it can be said that 0.1 unit increase in risk would increase the log of citations count by 0.667, keeping other variables constant. For example, if the current citations count for a firm is 100 and the risk is increased by 0.1 unit, then the new citations count would be equal to $100 * e^{0.667}$ or 194.83 citations count. Therefore, we can say that coefficient of risk is statistically as well as economically significant.

As to our control variables, we observe several significant relations that are consistent with Chkir et al. (2021). In particular we find that large firms with higher cash tend to engage in more innovative activities. The result is statistically significant at 1% level. We also find that plant, property and equipment is negatively related to innovation and is statistically significant at the 1% level. In addition, leverage is negatively (at the 1% level) related to innovation. Finally, dividend is negatively associated with innovation and statistically significant at the 1% level.

3.2 Robustness Tests

To assess the robustness of the model, we implement the Poisson model and negative binomial model using two additional measures of risk Risk2 and Risk3 as defined in Equation 3.3 and Equation 3.4 respectively. The dependent variable is the count of citations and the control variables are also the same except for one. In risk measure 2, return is defined as EBIT (Earnings before interest and taxes) divided by the sales of the firm.The detailed results of the Poisson regression using Risk2 and Risk3 is given in Table 3.6 and Table 3.8 respectively.

3.2.1 Risk Measure 2

Here the proxy of risk-taking is the volatility of return on sales over a rolling period of 4 years. Here the return is defined as the ratio of earnings before interest and taxes (EBIT) and Sales.

The risk measure 2 is defined as:

$$Risk2 = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}(E_{i,t} - \frac{1}{T}\sum_{t=1}^{T}E_{i,t})^2 | T = 4} \tag{3.3}$$

Where $E_{i,t} = \frac{EBIT_{i,t}}{Sales_{i,t}}$

$EBIT_{i,t}$ is defined as the earnings before interest and taxes for firm i in year t.

$Sales_{i,t}$ is defined as the sales for firm i in year t; T over (+1 to +4, +2 to +5, +3 to +6 ...)

3.2.2 Risk Measure 3

Here the proxy of risk-taking is the difference of maximum ROA and the minimum ROA over a period of four years. Here the return on assets is defined as the ratio of earnings before interest and taxes (EBIT) and total assets (At).

The risk measure 3 is defined as:

$$Risk3 = Max(E_{i,t}) - Min(E_{i,t}) \tag{3.4}$$

Where $E_{i,t}$= $EBIT_{i,t}/At_{i,t}$

$EBIT_{i,t}$ is defined as the earnings before interest and taxes for firm i in year t.

$At_{i,t}$ is defined as the total assets for firm i in year t; T over (+1 to +4, +2 to +5, ...).

The results are similar to the Poisson regression model using Risk measure 1. The dispersion statistic for both models is as follows:

1. Poisson Model using Risk measure 2

$$dispersion\ statistic = \frac{Pearson\ Chi2\ statistic}{residual\ degrees\ of\ freedom} = \frac{28400000}{1366} = 20790.62$$

2. Poisson Model using Risk measure 3

$$dispersion\ statistic = \frac{Pearson\ Chi2\ statistic}{residual\ degrees\ of\ freedom} = \frac{29300000}{1366} = 21449.48$$

Both Poisson models are highly over dispersed as the dispersion statistic is much greater than 1.0. To overcome the problem of over dispersion, negative binomial model is implemented. The value of α for which the dispersion statistic is closest to 1 is 2.35 for model using Risk measure 2 and 2.5 for model using Risk measure 3. The detailed summary of the negative binomial regression is shown in Table 3.7 and Table 3.9.The dispersion statistic for both models is as follows:

1. Negative Binomial Model using Risk measure 2 (at $\alpha = 2.35$)

$$dispersion\ statistic = \frac{Pearson\ Chi2\ statistic}{residual\ degrees\ of\ freedom} = \frac{1380}{1366} = 1.01$$

2. Negative Binomial Model using Risk measure 3 (at $\alpha = 2.5$)

$$dispersion\ statistic = \frac{Pearson\ Chi2\ statistic}{residual\ degrees\ of\ freedom} = \frac{1370}{1366} = 1.0029$$

Table 3.6 – The impact of Risk2 on innovation using Poisson Model

This table represents the results of the Poisson regression for risk measure 2. The dependent variable is the citation counts received by the firm on its published patents.The risk measure, risk2, corresponds to the volatility of Return computed over a rolling period of 4 years.Here Return is defined as the EBIT(Earnings before Interest and Taxes)/ Sales. The control variables are size (defined as log of total assets), Plant Property and Equipment (defined as PPE/total assets), Leverage (defined as total debt/total assets), Capital Expenditures (defined as capital expenditures/total assets), Cash (defined as cash/total assets), Dividend (defined as dividend/total assets). Our sample consists of the data on private firms. The citations data is obtained from the Derwent Innovation Database. The control variables are obtained from the compustat database. The data is from 2007 to 2015. To compute the standard deviation over a rolling period of 4 years only the companies that have more than 7 years of data are kept. The citations data and the data on control variables are merged in the final table resulting into 1375 firm-year observation

Df Residuals:	1366	**Pearson chi2:**	2.84e+07

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	2.6657	0.003	975.062	0.000	2.660	2.671
Risk2	0.6533	0.008	87.007	0.000	0.639	0.668
size	0.7450	0.000	3013.058	0.000	0.745	0.745
Return	-0.5068	0.003	-196.152	0.000	-0.512	-0.502
Leverage	-3.1309	0.002	-1426.387	0.000	-3.135	-3.127
Plant Property & Equipment	-6.8867	0.004	-1893.821	0.000	-6.894	-6.880
Capital Expenditures	18.7206	0.012	1517.257	0.000	18.696	18.745
Cash	1.5619	0.004	438.155	0.000	1.555	1.569
Dividend	-0.1707	0.010	-17.290	0.000	-0.190	-0.151

Table 3.7 – The impact of Risk2 on innovation using NB2 Model

This table represents the results of the Negative Binomial regression for risk measure 2 at $\alpha = 2.35$. The dependent variable is the citation counts received by the firm on its published patents.The risk measure, risk2, corresponds to the volatility of Return computed over a rolling period of 4 years.Here Return is defined as the EBIT(Earnings before Interest and Taxes)/ Sales. The control variables are size (log of total assets), Plant Property and Equipment (PPE/total assets), Leverage (total debt/total assets), Capital Expenditures (capital expenditures/total assets), Cash (cash/total assets), Dividend (dividend/total assets). Our sample consists of the data on private firms. The citations data is obtained from the Derwent Innovation Database. The control variables are obtained from the compustat database. The data is from 2007 to 2015. To compute the standard deviation over a rolling period of 4 years only the companies that have more than 7 years of data are kept. The citations data and the data on control variables are merged in the final table resulting into 1375 firm-year observations.

Df Residuals:	1366	**Pearson chi2:**	1.38e+03			

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	2.9326	0.344	8.514	0.000	2.257	3.608
Risk2	2.9734	0.913	3.258	0.001	1.185	4.762
size	0.6719	0.033	20.458	0.000	0.608	0.736
Return	-0.8911	0.358	-2.490	0.013	-1.592	-0.190
Leverage	-1.8619	0.278	-6.692	0.000	-2.407	-1.317
Plant Property & Equipment	-4.5681	0.446	-10.237	0.000	-5.443	-3.694
Capital Expenditures	12.6940	1.823	6.962	0.000	9.120	16.268
Cash	1.3714	0.508	2.702	0.007	0.376	2.366
Dividend	1.4090	1.453	0.970	0.332	-1.439	4.257

Table 3.8 – The impact of Risk3 on innovation using Poisson Model

This table represents the results of the Poisson regression for risk measure 3. The dependent variable is the citation counts received by the firm on its published patents. The risk measure, risk3, corresponds to the difference between the maximum and minimum value of ROA(return on Assets) computed over a rolling period of 4 years.Here ROA is defined as the EBIT(Earnings before Interest and Taxes)/ At(total assets). The control variables are size (defined as log of total assets), return on assets (defined as Earnings before interest and taxes/total assets), Plant Property and Equipment (defined as PPE/total assets), Leverage (defined as total debt/total assets), Capital Expenditures (defined as capital expenditures/total assets), Cash (defined as cash/total assets), Dividend (defined as dividend/total assets). Our sample consists of the data on private firms. The citations data is obtained from the Derwent Innovation Database. The control variables are obtained from the compustat database. The data is from 2007 to 2015. To compute the measure over a rolling period of 4 years only the companies that have more than 7 years of data are kept. The citations data and the data on control variables are merged in the final table resulting into 1375 firm-year observations.

Df Residuals:	1366	**Pearson chi2:**	2.93e+07

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	2.4219	0.003	872.927	0.000	2.416	2.427
Risk3	0.3023	0.005	60.079	0.000	0.292	0.312
size	0.7359	0.000	3043.839	0.000	0.735	0.736
Return on Assets	1.3774	0.004	312.149	0.000	1.369	1.386
Leverage	-2.7188	0.002	-1197.432	0.000	-2.723	-2.714
Plant Property & Equipment	-6.6431	0.004	-1857.339	0.000	-6.650	-6.636
Capital Expenditures	17.5924	0.013	1379.918	0.000	17.567	17.617
Cash	1.5853	0.004	451.156	0.000	1.578	1.592
Dividend	-1.9337	0.012	-162.297	0.000	-1.957	-1.910

Table 3.9 – The impact of Risk3 on innovation using NB2 Model

This table represents the results of the Negative Binomial regression for risk measure 3 at $\alpha = 2.5$. The dependent variable is the citation counts received by the firm on its published patents.The risk measure, risk3, corresponds to the difference between the maximum and minimum value of ROA(return on Assets) computed over a rolling period of 4 years.Here ROA is defined as the EBIT(Earnings before Interest and Taxes)/ At(total assets). The control variables are size (log of total assets), Plant Property and Equipment (PPE/total assets), Leverage (total debt/total assets), Capital Expenditures (capital expenditures/total assets), Cash (cash/total assets), Dividend (dividend/total assets). Our sample consists of the data on private firms. The citations data is obtained from the Derwent Innovation Database. The control variables are obtained from the compustat database. The data is from 2007 to 2015. To compute the measure over a rolling period of 4 years only the companies that have more than 7 years of data are kept. The citations data and the data on control variables are merged in the final table resulting into 1375 firm-year observations.

Df Residuals:	1366	**Pearson chi2:**	1.37e+03			
	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	2.7911	0.362	7.703	0.000	2.081	3.501
Risk3	2.9783	0.786	3.791	0.000	1.438	4.518
size	0.6674	0.034	19.866	0.000	0.602	0.733
Return on Assets	-0.0595	0.577	-0.103	0.918	-1.190	1.071
Leverage	-1.9319	0.286	-6.749	0.000	-2.493	-1.371
Plant Property & Equipment	-3.8477	0.453	-8.489	0.000	-4.736	-2.959
Capital Expenditures	10.9049	1.925	5.666	0.000	7.133	14.677
Cash	1.4044	0.521	2.697	0.007	0.384	2.425
Dividend	-2.0795	1.534	-1.356	0.175	-5.086	0.927

Table 3.7 and Table 3.9 corroborates our findings reported in Table 3.4. Although the values of the coefficient of risk 2 and risk 3 are significant, we can see that the magnitude of the coefficient has decreased when compared to coefficient of risk 1. It implies that the innovation activities are less impacted if risk 2 and risk 3 are taken as a proxy for risk. For every 0.1 unit increase in risk 2 and risk 3, the log of citations count increases by 0.29734 and 0.29783 respectively.

Section 4

Collaboration and Innovation

The second part of our research focuses on exploring the effect of collaborative activities of the firm on innovation done by the firm (measured by the citations count). Privatization is associated with an increase in collaboration (Somé et al. (2021)). We extend this analysis by exploring the effect of collaboration on the innovation activities of the firm. The dependent variable is the citations count received by the firm on its published patents.The detailed description of the data collection on citations count is given in chapter 1.

Since the dependent variable is still the citations count, we applied the Poisson and the negative binomial models as discussed in chapter 2. The detailed summary for the Poisson regression is given in Table 4.1.

The Poisson model is highly over dispersed as we can see from the dispersion statistic:

$dispersion\ statistic = \frac{Pearson\ Chi2\ statistic}{residual\ degrees\ of\ freedom} = \frac{40200000}{2104} = 19106.46$

The dispersion statistic is 19106.46 which is much greater than 1.0. To overcome this problem we implemented the negative binomial model. Since in a negative binomial model, the parameter α needs to be inserted as a constant, the value of α for which the model's dispersion statistic is closest to 1.0 is 2.775. The detailed summary of the negative binomial regression is shown in Table 4.2. This model is properly specified as we can see from the dispersion statistic:

$dispersion\ statistic = \frac{Pearson\ Chi2\ statistic}{residual\ degrees\ of\ freedom} = \frac{2190}{2104} = 1.04$

Table 4.1 – The impact of Collaboration 1 on innovation using Poisson Model

This table represents the results of the Poisson regression for collaboration measure 1. The dependent variable is the citation counts received by the firm on its published patents. The collaboration measure is the average of inventor count. The control variables are size (defined as log of total assets), return on assets (defined as Earnings before interest and taxes/total assets), Plant Property and Equipment (defined as PPE/total assets), Leverage (defined as total debt/total assets), Capital Expenditures (defined as capital expenditures/total assets), Cash (defined as cash/total assets), Dividend (defined as dividend/total assets). Our sample consists of the data on private firms. The citations count data and the inventor count data is obtained from the Derwent Innovation Database. The control variables are obtained from the compustat database. The data is from 2007 to 2015. The citations data, inventor count data and the data on control variables are merged in the final table resulting into 2113 firm-year observations.

Df Residuals:	2104	**Pearson chi2:**	4.02e+07			
	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	2.8058	0.002	1178.410	0.000	2.801	2.810
Average of Inventor Count	-0.2775	0.000	-757.291	0.000	-0.278	-0.277
size	0.7551	0.000	3643.783	0.000	0.755	0.756
Return on Assets	2.4540	0.004	647.495	0.000	2.447	2.461
Plant Property & Equipment	-6.3091	0.003	-2059.006	0.000	-6.315	-6.303
Leverage	-2.7535	0.002	-1446.839	0.000	-2.757	-2.750
Capital Expenditures	15.9899	0.012	1343.227	0.000	15.967	16.013
Cash	1.6548	0.003	531.575	0.000	1.649	1.661
Dividend	-3.9265	0.011	-360.779	0.000	-3.948	-3.905

Table 4.2 – The impact of Collaboration 1 on innovation using NB2 Model

This table represents the results of the negative binomial regression for collaboration measure 1 at $\alpha = 2.775$. The dependent variable is the citation counts received by the firm on its published patents. The collaboration measure is the average of inventor count. The control variables are size (defined as log of total assets), return on assets (defined as Earnings before interest and taxes/total assets), Plant Property and Equipment (defined as PPE/total assets), Leverage (defined as total debt/total assets), Capital Expenditures (defined as capital expenditures/total assets), Cash (defined as cash/total assets), Dividend (defined as dividend/total assets). Our sample consists of the data on private firms. The citations count data and the inventor count data is obtained from the Derwent Innovation Database. The control variables are obtained from the compustat database. The data is from 2007 to 2015. The citations data, inventor count data and the data on control variables are merged in the final table resulting into 2113 firm-year observations.

Df Residuals:	2104	**Pearson chi2:**	2.19e+03			
	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	3.7414	0.314	11.919	0.000	3.126	4.357
Average of Inventor Count	-0.2073	0.048	-4.307	0.000	-0.302	-0.113
size	0.6323	0.028	22.456	0.000	0.577	0.688
Return on Assets	-0.1620	0.504	-0.321	0.748	-1.150	0.826
Plant Property & Equipment	-4.1804	0.377	-11.101	0.000	-4.919	-3.442
Leverage	-1.8827	0.244	-7.722	0.000	-2.361	-1.405
Capital Expenditures	11.7707	1.739	6.769	0.000	8.362	15.179
Cash	1.8597	0.450	4.130	0.000	0.977	2.742
Dividend	-1.3063	1.188	-1.099	0.272	-3.635	1.022

Results in Table 4.2 show the evidence of the effects of collaboration on firms' innovative activities. A firm's collaboration and innovation are negatively associated. The coefficient is significant at 1 % level. This finding is opposite to hypothesis 2. One plausible explanation is the geographic distribution of our sample. As shown in Table 1.2, our sample has 64.23% U.S. companies. The empirical study of the effects of collaboration on innovation is done in Inoue and Liu (2015).The authors study innovation patterns in Japanese and U.S. firms and find out a negative relationship between collaboration and innovation in the U.S. firms. According to Inoue and Liu (2015), typically, U.S. workers are subject to the strong pressure/incentive for the immediate result, implying that taking time for U.S. inventors to deepen their collaborations is not a good strategy. This lack of time leads to poor collaboration making the patents less impact full. Our proxy for innovation is citations count which measures the impact of patents filed by the firm. Hence, due to poor collaboration, we are observing a negative relationship between collaboration and innovation activities of the firm.

The result in Table 4.2 shows us that the coefficient of collaboration has less impact on the citations count. Although, the coefficient is statistically significant, the impact of collaboration, as measured by average of inventor count, is not significant. For every 1 unit increase in the average of inventor count, log of citations count decrease by 0.2073. In other words, if the citations count for a firm is 100 and the average of inventor count on each patent increases by 1, then new citations count would be equal to $100 * e^{-0.2073}$ or 81.27 citations count. Hence, we can say that the coefficient for collaboration is not economically significant.

We observe the same results for our control variables that are consistent with Chkir et al. (2021). In particular we find that large firms with higher cash tend to engage in more innovative activities. The result is statistically significant at 1% level. We also find that plant, property and equipment is negatively related to innovation and is statistically significant at the 1% level. In addition, leverage is negatively (at the 1% level) related to innovation. Finally, dividend is negatively associated with innovation and statistically significant at the 1% level.

Robustness check:

In order to conduct the robustness check, we performed the analysis using the average of the assignee count as the measure of collaboration. The detailed results of the Poisson regression and the negative binomial regression are given in Table 4.3 and Table 4.4 respectively.

Table 4.3 – The impact of Collaboration 2 on innovation using Poisson Model

This table represents the results of the Poisson regression for collaboration measure 2. The dependent variable is the citation counts received by the firm on its published patents. The collaboration measure is the average of assignee count. The control variables are size (defined as log of total assets), return on assets (defined as Earnings before interest and taxes/total assets), Plant Property and Equipment (defined as PPE/total assets), Leverage (defined as total debt/total assets), Capital Expenditures (defined as capital expenditures/total assets), Cash (defined as cash/total assets), Dividend (defined as dividend/total assets). Our sample consists of the data on private firms. The citations count data and the assignee count data is obtained from the Derwent Innovation Database. The control variables are obtained from the compustat database. The data is from 2007 to 2015. The citations data, assignee count data and the data on control variables are merged in the final table resulting into 2113 firm-year observations.

Df Residuals:	2104	**Pearson chi2:**	4.17e+07

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	2.8581	0.002	1183.990	0.000	2.853	2.863
Average of Assignee Count	-0.3256	0.001	-561.339	0.000	-0.327	-0.324
size	0.7289	0.000	3623.481	0.000	0.729	0.729
Return on Assets	1.4520	0.004	414.006	0.000	1.445	1.459
Plant Property & Equipment	-6.5030	0.003	-2153.704	0.000	-6.509	-6.497
Leverage	-2.8097	0.002	-1486.772	0.000	-2.813	-2.806
Capital Expenditures	18.0326	0.011	1598.960	0.000	18.011	18.055
Cash	1.6101	0.003	524.197	0.000	1.604	1.616
Dividend	-3.6342	0.011	-339.676	0.000	-3.655	-3.613

Table 4.4 – The impact of Collaboration 2 on innovation using NB2 Model

This table represents the results of the negative binomial regression for collaboration measure 2 at $\alpha = 3.0$. The dependent variable is the citation counts received by the firm on its published patents. The collaboration measure is the average of assignee count. The control variables are size (defined as log of total assets), return on assets (defined as Earnings before interest and taxes/total assets), Plant Property and Equipment (defined as PPE/total assets), Leverage (defined as total debt/total assets), Capital Expenditures (defined as capital expenditures/total assets), Cash (defined as cash/total assets), Dividend (defined as dividend/total assets). Our sample consists of the data on private firms. The citations count data and the assignee count data is obtained from the Derwent Innovation Database. The control variables are obtained from the compustat database. The data is from 2007 to 2015. The citations data, assignee count data and the data on control variables are merged in the final table resulting into 2113 firm-year observations.

Df Residuals:	2104	**Pearson chi2:**	2.13e+03			
	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	3.4575	0.332	10.410	0.000	2.807	4.108
Average of Assignee Count	-0.2653	0.090	-2.949	0.003	-0.442	-0.089
size	0.6367	0.029	21.946	0.000	0.580	0.694
Return on Assets	-0.5066	0.520	-0.974	0.330	-1.526	0.513
Plant Property & Equipment	-4.2428	0.390	-10.869	0.000	-5.008	-3.478
Leverage	-1.7987	0.254	-7.093	0.000	-2.296	-1.302
Capital Expenditures	12.8263	1.785	7.188	0.000	9.329	16.324
Cash	2.1323	0.468	4.554	0.000	1.215	3.050
Dividend	-1.4894	1.235	-1.206	0.228	-3.910	0.932

The robustness measure for collaboration, average of assignee count, is statistically significant at 1% level. The measure validates our results from Table 4.2 and we can say that collaboration has a negative effect on innovation activities of a firm. Although, the coefficient of average of assignee count is statistically significant, the magnitude of the coefficient is small and we can say that it not economically significant. If the citations count for a firm is 100 and the average of assignee count on each patent increases by 1, then new citations count would be equal to $100 * e^{-0.2653}$ or 76.69 citations count. Hence, we can say that the coefficient is not economically significant.

Section 5

Ownership and Innovation

The third part of our research focuses on exploring the effect of state ownership of the firm on innovation done by the firm (measured by the citations count). State ownership is negatively related to risk taken by a firm (Boubakri et al. (2013)). We extend this analysis by exploring the effect of state ownership on the innovation activities of the firm. The dependent variable is the citations count received by the firm on its published patents. The detailed description of the data collection on citations count is given in chapter 1.

Since the dependent variable is still the citations count, we applied the Poisson and the negative binomial models as discussed in chapter 2. The detailed summary for the Poisson regression is given in Table 5.1.

The Poisson model is highly over dispersed as we can see from the dispersion statistic:

$$dispersion\ statistic = \frac{Pearson\ Chi2\ statistic}{residual\ degrees\ of\ freedom} = \frac{39400000}{2104} = 18726.23$$

The dispersion statistic is 18726.23 which is much greater than 1.0. To overcome this problem we implemented the negative binomial model. The value of α for which the model's dispersion statistic is closest to 1.0 is 2.8. The detailed summary of the negative binomial regression is shown in Table 5.2. This model is properly specified as we can see from the dispersion statistic:

$$dispersion\ statistic = \frac{Pearson\ Chi2\ statistic}{residual\ degrees\ of\ freedom} = \frac{2110}{2104} = 1.002$$

Table 5.1 – The impact of State Ownership on innovation using Poisson Model

This table represents the results of the Poisson regression for state ownership on innovation. The dependent variable is the citation counts received by the firm on its published patents. The ownership measure is state1stclass which is the percentage of first class shares owned by state owned enterprise(SOE). The control variables are size (defined as log of total assets), return on assets (defined as Earnings before interest and taxes/total assets), Plant Property and Equipment (defined as PPE/total assets), Leverage (defined as total debt/total assets), Capital Expenditures (defined as capital expenditures/total assets), Cash (defined as cash/total assets), Dividend (defined as dividend/total assets). Our sample consists of the data on private firms. The citations count data is obtained from the Derwent Innovation Database and the ownership data is obatined from Narjess Boubakri and used in Boubakri et al. (2013). The control variables are obtained from the compustat database. The data is from 2007 to 2015. The citations data, state ownership data and the data on control variables are merged in the final table resulting into 2113 firm-year observations.

Df Residuals:	2104	**Pearson chi2:**	3.94e+07			
	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	2.3201	0.002	989.295	0.000	2.315	2.325
state1stclass	-4.4195	0.007	-642.859	0.000	-4.433	-4.406
size	0.7327	0.000	3638.171	0.000	0.732	0.733
Return on Assets	1.5843	0.004	443.462	0.000	1.577	1.591
Plant Property & Equipment	-5.7843	0.003	-1888.354	0.000	-5.790	-5.778
Leverage	-2.7536	0.002	-1461.203	0.000	-2.757	-2.750
Capital Expenditures	16.9386	0.011	1492.806	0.000	16.916	16.961
Cash	1.6519	0.003	536.169	0.000	1.646	1.658
Dividend	-4.7458	0.011	-421.060	0.000	-4.768	-4.724

Table 5.2 – The impact of State Ownership on innovation using NB2 Model

This table represents the results of the negative binomial regression for state ownership on innovation at $\alpha = 2.8$. The dependent variable is the citation counts received by the firm on its published patents. The ownership measure is state1stclass which is the percentage of first class shares owned by state owned enterprise(SOE). The control variables are size (defined as log of total assets), return on assets (defined as Earnings before interest and taxes/total assets), Plant Property and Equipment (defined as PPE/total assets), Leverage (defined as total debt/total assets), Capital Expenditures (defined as capital expenditures/total assets), Cash (defined as cash/total assets), Dividend (defined as dividend/total assets). Our sample consists of the data on private firms. The citations count data is obtained from the Derwent Innovation Database and the ownership data is obatined from Narjess Boubakri and used in Boubakri et al.(2013). The control variables are obtained from the compustat database. The data is from 2007 to 2015. The citations data, state ownership data and the data on control variables are merged in the final table resulting into 2113 firm-year observations.

Df Residuals:	2104	**Pearson chi2:**	2.11e+03			
	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	2.8599	0.305	9.372	0.000	2.262	3.458
state1stclass	-4.8812	0.441	-11.061	0.000	-5.746	-4.016
size	0.6536	0.028	23.212	0.000	0.598	0.709
Return on Assets	-0.3853	0.502	-0.767	0.443	-1.369	0.599
Plant Property & Equipment	-3.7618	0.381	-9.879	0.000	-4.508	-3.015
Leverage	-1.8030	0.245	-7.361	0.000	-2.283	-1.323
Capital Expenditures	13.2209	1.726	7.659	0.000	9.837	16.604
Cash	2.0239	0.452	4.474	0.000	1.137	2.910
Dividend	-1.6866	1.193	-1.414	0.157	-4.025	0.651

Results in Table 5.2 show the evidence of the effects of state ownership on firms' innovation activities. The results lend empirical support to the view that state ownership and corporate innovation are negatively related. This association is statistically significant at 1% level. In line with the expectations, these results support hypothesis 3, which posits that state ownership and corporate innovation are negatively associated. Although, the results are statistically significant, the coefficient is not economically significant. For 1% or 0.01 unit increase in state ownership, the log of citations count is decreased by 0.0488. If the firm has 100 citations count and the state ownership is increased by 0.01, the new citations count is $100 * e^{-0.0488}$ or 95.23 citations count. Hence, we can say that the impact of state ownership on citations count of a firm is not meaningful.

The coefficients and significance of control variables reaffirm our previous results. We find that large firms with higher cash tend to engage in more innovative activities. The result is statistically significant at 1% level.We also find that plant, property and equipment is negatively related to innovation and is statistically significant at the 1% level. In addition, leverage is negatively (at the 1%level) related to innovation. Finally, dividend is negatively associated with innovation and statistically significant at the 1% level.

Table 5.3 – The impact of Risk, Collaboration and Ownership on innovation

This table represents the results of the negative binomial regression for risk, collaboration and state ownership on innovation. The dependent variable is the citation counts received by the firm on its published patents.The risk measure, risk1, corresponds to the volatility of ROA(return on Assets) computed over a rolling period of 4 years. Here ROA is defined as the EBIT/At(total assets). The collaboration measure is the average of inventor count. The ownership measure is state1stclass which is the percentage of first class shares owned by state owned enterprise(SOE). The control variables are size (defined as log of total assets), return on assets (defined as Earnings before interest and taxes/total assets), Plant Property and Equipment (defined as PPE/total assets), Leverage (defined as total debt/total assets), Capital Expenditures (defined as capital expenditures/total assets), Cash (defined as cash/total assets), Dividend (defined as dividend/total assets). Our sample consists of the data on private firms. The citations count data is obtained from the Derwent Innovation Database and the ownership data is obtained from Narjess Boubakri and used in Boubakri et al.(2013). The control variables are obtained from the compustat database. The data is from 2007 to 2015. The citations data, risk measure, collaboration, state ownership data and the data on control variables are merged in a final table resulting into 1375 firm-year observations.

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	3.0090	0.360	8.367	0.000	2.304	3.714
Risk1	6.0726	1.667	3.642	0.000	2.804	9.341
Average of Inventor Count	-0.1361	0.055	-2.469	0.014	-0.244	-0.028
state1stclass	-4.9566	0.483	-10.266	0.000	-5.903	-4.010
size	0.6848	0.032	21.076	0.000	0.621	0.748
Return on Assets	0.2088	0.553	0.377	0.706	-0.876	1.293
Plant Property & Equipment	-3.4425	0.436	-7.893	0.000	-4.297	-2.588
Leverage	-1.9529	0.272	-7.187	0.000	-2.485	-1.420
Capital Expenditures	10.2100	1.850	5.519	0.000	6.584	13.836
Cash	1.0761	0.494	2.177	0.029	0.107	2.045
Dividend	-2.4337	1.455	-1.672	0.094	-5.286	0.419

Limitations

Although our results are statistically as well as economically significant, there are some concerns about endogeneity that needs to be explored further. Specifically, the omitted variable bias and the reverse causality problem needs to be addressed to examine whether our findings are driven by endogeneity problems. To do so, we can implement a two-stage least squares (2SLS) approach. The one-year lagged values of our risk measures (i.e., risk1, risk2, and risk3) can be used as instruments. In the first stage of the 2SLS, we regress the risk measures against their instruments along with the control variables. In the second stage, we use the fitted values from the stage 1 as explanatory variables. If the results still gives a positive relationship between risk and innovation, we can say that our results are not driven by omitted variable bias.

With regard to the reverse causality problem, it is reasonable to think that increased innovation makes the firm more risky. To disentangle this issue, we can perform an instrumental generalized method of moments using the same instruments mentioned above. If the coefficients are still significant, we can say that our results are unlikely to suffer from endogeneity issues.

Conclusion

In this study, we rely on a unique database of 246 privatized firms from 19 countries to investigate the impact of risk-taking on corporate innovation, where we measure innovation using the citations count and risk-taking using the volatility of earnings over four overlapping periods following the divestitures of SOEs. Additionally, we also investigate the impact of collaboration and state ownership on corporate innovation.

We argue that the shift in ownership and control that accompanies privatization will affect corporate innovation activities. Post privatization companies will tend to be more commercially oriented rather than broader national interest. In order to have competitive edge in the market, companies that are more risk-oriented would invest into new technologies. Additionally, we argue that companies' collaboration activities help companies to remain competitive in today's environment; by collaborating, a firm can increase its innovation capabilities. Also, we hypothesized that state ownership is negatively related to corporate innovation.

In a negative binomial regression that controls for firm-level variables, we provide evidence that risk-taking is positively related to corporate innovation. Moreover, we found a significant negative relation of state ownership and collaboration on innovation.

Bibliography

Adler, K., Ahn, J., Dao, M. C., and Cubeddu, L. M. (18 Jan. 2019). Innovation and corporate cash holdings in the era of globalization. *IMF Working Papers*, 2019(017):A001.

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

Boubakri, N. and Cosset, J.-C. (1998). The financial and operating performance of newly privatized firms: Evidence from developing countries. *The Journal of Finance*, 53(3):1081–1110.

Boubakri, N., Cosset, J.-C., and Saffar, W. (2013). The role of state and foreign owners in corporate risk-taking: Evidence from privatization. *Journal of Financial Economics*, 108(3):641–658.

Boutchkova, M. K. and Megginson, W. L. (2000). Privatization and the rise of global capital markets. *Financial Management*, 29(4):31–75.

Cameron, A. C. and Trivedi, P. K. (2013). *Regression analysis of count data*, volume 53. Cambridge university press.

Chang, X., Fu, K., Low, A., and Zhang, W. (2015). Non-executive employee stock options and corporate innovation. *Journal of Financial Economics*, 115(1):168–188.

Chkir, I., El Haj Hassan, B., Rjiba, H., and Saadi, S. (2021). Does corporate social responsibility influence corporate innovation? international evidence. *Emerging Markets Review*, 46:100746. Corporate Social Responsibility in Emerging Markets.

Cho, C., Halford, J. T., Hsu, S., and Ng, L. (2016). Do managers matter for corporate innovation? *Journal of Corporate Finance*, 36:206–229.

Djankov, S. and Murrell, P. (2002). Enterprise restructuring in transition: A quantitative survey. *Journal of Economic Literature*, 40(3):739–792.

Estrin, S., Hanousek, J., Kočenda, E., and Svejnar, J. (2009). The effects of privatization and ownership in transition economies. *Journal of Economic Literature*, 47(3):699–728.

Hilbe, J. M. (2014). *Modeling count data*. Cambridge University Press.

Inoue, H. and Liu, Y.-Y. (2015). Revealing the intricate effect of collaboration on innovation. *PLOS ONE*, 10(3):1–16.

Munari, F., Roberts, E. B., and Sobrero, M. (2002). Privatization processes and the redefinition of corporate rd boundaries. *Research Policy*, 31(1):31–53.

Okten, C. and Arin, K. P. (2006). The effects of privatization on efficiency: How does privatization work? *World Development*, 34(9):1537–1556.

Somé, H. Y., Cano-Kollmann, M., Mudambi, R., and Cosset, J.-C. (2021). The effect of privatization on the characteristics of innovation. *Financial Management*.

Symeonidis, G. (1996). Innovation, firm size and market structure. (161).

Tan, Y., Tian, X., Zhang, X., and Zhao, H. (2020). The real effect of partial privatization on corporate innovation: Evidence from china's split share structure reform. *Journal of Corporate Finance*, 64:101661.