The Report Committee for Xue Gao
Certifies that this is the approved version of the following report:

The effect of government subsidies on firms' private investment and innovation capacity: Evidence from the firm-level data in the Chinese wind industry

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor: ______________________________
Michael J. Daniels

____
Timothy Keitt

The effect of government subsidies on firms' private investment and innovation capacity: Evidence from the firm-level data in the Chinese wind industry

by

Xue Gao, B ECO; M ECO

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Statistics

The University of Texas at Austin

May 2017

Abstract

The effect of government subsidies on firms' private investment and innovation capacity: Evidence from the firm-level data in the Chinese wind industry

Xue Gao, M. S. STAT.
The University of Texas at Austin, 2017

Supervisor: Michael J. Daniels

Abstract: Scientific knowledge is a public good, so a firm's investment in technology innovation is below the most socially efficient level. To correct this market failure, government subsidies are necessary to stimulate firms' private investments and innovation activities, especially at the early development stage of an industry. Although there is a great deal of research studying the relationship between government subsidies and innovations in developed countries, only a few studies focus on developing countries, and even fewer focus on emerging industries in developing countries. This paper aims at studying the effect of the Chinese government's subsidies on firms' private investments and innovation outcomes by using Chinese wind firms' data in 2007. In order to capture the average treatment effect of government subsidies, this paper used the propensity score matching method. The results show that government subsidies have no effect on firms' innovation outcomes and private R&D investments. This paper further suggests that these non-ideal policy outcomes might result from the immaturity of the Chinese wind industry in 2007, the insufficient amount of government subsidies and/or the uncertainty of subsidies' persistence.

Table of Contents

List of Tables

List of Figures

Chapter 1: *Introduction*

Climate change is an increasingly serious problem in the world. In 2007, the United Nations Intergovernmental Panel on Climate Change (IPCC) released the fourth assessment report on climate change, which said that most of the observed increase in global average temperatures since the mid-20th century is very likely caused by humans (IPCC, 2008). IPCC's fifth assessment report reaffirmed the conclusion that global warming is occurring and is caused by humans (IPCC, 2014). Based on these scientific determinations of climate change, low carbon development is necessary for our world to be sustainable. The biggest concern about carbon reduction comes from the very possible negative impact on economic growth. To balance the economic growth and carbon reduction, promoting renewable energy technology innovation is an important component of the ideal path. Technology innovation in the clean energy industry might provide new economic growth opportunities and reduce carbon emissions as well. Such new economic growth opportunities are more important to developing countries when they face the pressure of carbon reduction.

Knowledge is a public good, which has positive externalities, so a firm's Research and Development (R&D) investment is below the most efficient level (Nelson, 1959; Arrow, 1962; Griliches, 1992; Hall, 2002). Technology innovation policies are necessary to correct such market failures. The goal of technology innovation policy is to influence or change the speed, direction and scale of technological innovation. Generally, there are three types of policy instruments, the supply-side policy instrument, the demand-side policy instrument and the environmental support policy instrument (Rothwell, 1985). Public funding (government subsidies) is a type of supply-side policy instrument that aims to help increase firms' incentives to undertake long-term technology innovation. The effect of public funding on firms' innovation is very consistent. A great deal of research has verified the importance of government funding in promoting firms' innovation. (Romer, 1989; Grossman and Helpman, 1993; Aghion and Howitt, 1990; Leyden and Link, 1991; Kole and Mulherin, 1997; Maryann, 2006; Nola Hewitt Dundas, 2008; Choi et al., 2011; etc.). But when it comes to the study of the effect of public investment on business investment, results are not consistent. Generally, public funding has two effects on private investments (business investment): public funding can stimulate private investments or it can crowd out private investments. The different results may be due to different types of subsidies, development stages of industry, characteristics of firms, or subsidy amounts (Tommy, 2009; Gullec et al., 2000; Duguet, 2004).

There is plenty of research studying the relationship between R&D investment and innovation in developed countries. But only a few studies focus on developing countries, and even fewer focus on emerging industries in developing countries. As the largest economy and biggest carbon emitter in the world, China's low carbon development plays a significant role in the global governance of climate change. China enacted the Renewable Energy Law in 2005. After 2005, China has been increasing government support for the renewable energy industry, with the aim of improving the manufacturing and innovation capacity of the renewable energy industry. Taking the wind industry as an example, Chinese cumulative installation of wind power capacity accounted for more than 30% of total global installation in 2014, which increased from only 1.7% in 2001. Beside the production capacity, the Chinese government also cares about the independent innovation capacity of the renewable energy industry, because innovation is the key to the sustainable development of emerging industries. Therefore, this study aims at evaluating the effect of Chinese government subsidies on firms' private investments and innovation outcomes.

To be specific, the purpose of the study is to answer whether Chinese government subsidies crowd out private investments and whether Chinese government subsidies have a positive effect on the firm's innovation capacity in the wind industry. The allocation of government subsidies is not random. Firms with large size or high technology innovation capacity are more likely to obtain government subsidies, so standard regression analyses cannot verify the causation between government subsidies and firms' private investments or firms' innovation capacity. Therefore, this study matches firms according to their main financial data and the number of firms' patents before 2007. By using matching, we can compare outcomes between firms obtaining government subsidies and their matched control firms that do not receive them. After matching, we should expect that obtaining government subsidies will be the main difference between each matching pair. Thus, we can assess the causal relationship between government subsidies and firms' private investment or innovation capacity.

Chapter 2: *Literature Review*

In the 1980s scholars began to study technological innovation from the perspective of system theory. Christopher Freeman is the first scholar who used the term "national innovation system." The economy of major western industrial countries declined very drastically in the 1980s, but the economy of Japan and several emerging countries in East Asia was very prosperous, which provided an exciting new topic for researchers. For example, Freeman was interested in factors that contributed to the economic growth in Japan, so he studied the role of the Japanese government (especially the Ministry of International Trade and Industry) and the Japanese strategy of "building the country by technology" in Japan's technological innovation mechanism, pointing out that the national innovation system contributed a lot to Japan's high-speed development.

Why it is so important to involve government and public policy in the innovation process? Technology innovation policy is a series of public policies that aim at influencing or changing the speed, direction or scale of technological innovation. The Organization for Economic Co-operation and Development (OECD) concluded that there were three reasons why government actions on technological innovation are important (OECD, 1999). The first reason is that technological innovation policies are necessary to correct market failure because knowledge is a public good, which has positive externalities. Since firms cannot gain all the benefits from their R&D investments and they cannot get any reimbursement for knowledge spillover, firms' R&D investments will be lower than the most socially efficient level. Many scholars have reached the same conclusion (Nelson, 1959; Arrow, 1962; Griliches, 1992; Hall, 2002). The second reason for government action is system failure. If system failure exists, it is difficult to deal with market failure efficiently. For example, if the innovation network and the system of knowledge diffusion are defective, then they will need government to fix the problems of organization management and speed up the technology diffusion. The third reason is to fulfill national strategies, such as energy independence and national defense strategy. Innovation policies should aim at reducing the uncertainty, learning cost and information asymmetry faced by technology inventors and technology adopters.

Although there are almost no difference of opinion on the general goal of technology innovation policy, when it comes to the specific technology innovation policy and how to choose a specific technology innovation policy tool, a variety of different views have emerged. Here I only focus on studies that are about public funding (including government R&D subsidies) and firms' innovation behaviors.

The emphasis of R&D investments in technology innovation comes from the endogenous growth model of technological progress in the 1980s. Griliehes (1980) firstly built the mathematical model to analyze the positive contribution of R&D investments, technology introduction and government funding on firms' productivity. Romer (1989), Grossman and Helpman (1993), and Aghion and Howitt (1990) further developed the above model and drew the conclusion that R&D activities are a source of firms' technological progress. Furman et al. (2002) found that the difference of R&D investments alone can explain 90% of the variation of innovation of OECD countries.

The results of the effect of public funding on firms' innovation are very consistent, which all verify the important role of government funding in promoting firms' innovation. But when it comes to the effect of public investments on private investments, there are always disagreements. Generally, public funding has two effects on private investments: public funding can stimulate private investments, but under some situations public funding can crowd out private investments and further weaken the effect of public funding on firm's innovation outcomes. Czarnitzki et al. (2007) collected 14 papers studying the relationship between government funding and firms' investments and only 2 papers confirmed the existence of significant crowd-out effect. Song Hong (2015) collected 19 papers from the top research journals and he found that 9 papers verified the existence of partial or full crowd-out effect. The existence of disagreements and inconsistent conclusions suggests that this research question needs further exploration.

The most common research design in verifying the effect of public funding on firms' innovation behaviors (including private investments and innovation outcomes) is the simple OLS regression. But the process of allocating government subsidies is not random. In order to take full advantage of public funding, governments always allocate government subsidies to the firms with high innovation capacity. Therefore, if we use government subsidies as the predictor variable, it will have the problem of endogeneity with firm's innovation capacity. Furthermore, older or larger firms or firms with riskier innovation programs are more willing to apply for government subsidies, so the problem of self-selection also exists (David et al., 2000; Klette et al., 2000). In order to better capture the causation between public funding and firms' innovation behaviors, the research design must consider the problem of endogeneity and selection bias.

Generally, there are four different research designs to address this problem in previous studies. The first way is to use an instrumental variable. The researchers look for a variable that is correlated with the subsidies (endogenous variable) but isn't correlated with measures of a firm's innovation performance. The instrumental variable used in Lichtenberg (1988) is the total contract value of government funding programs, which can address these potential problems under reasonable assumptions. Wallsten (2000) adapted the instrumental variable used by Lichtenberg (1988). He used the U.S. Small Business

Innovation Research (SBIR) program budget as his instrumental variable. Clausen, T. H., (2009) used two instrumental variables, including the total amount of public funding at the industry level from different government ministries and firms' distance to the headquarter of National Research Council in Oslo. These instrumental variables have limitations (either weak or invalid to some extent), so they also cause a lot of criticism.

The second research design is the Heckman selection model, which includes two equations. The first equation describes the relationship between the outcome variable and several covariates. The second equation describes the relationship between a dummy variable (e.g., firms with subsidies and firms without subsidies) and a vector of covariates. Then we need to estimate these two equations simultaneously. Busom (2000) used the Heckman selection model and found the existence of the partial crowd-out effect in Spain. One interesting finding in this paper is that small firms are more likely to obtain a subsidy than large firms in Spain. Hussinger (2003) used parametric and semi-parametric selection models to study German manufacturing. His results showed that the average treatment effect of public funding on firms' private R&D investment is positive. González et al. (2005) also used Spanish manufacturing firms. They found that the effect of government subsidies in stimulating private investment is very limited.

The third research design is the difference-in-difference model. Lach (2002) applied the difference in difference model to Israel manufacturing firms in the 1990s. He believed that the difference in difference model can address both problems of common observed covariates and time-invariance differences between firms with subsidies and firms without subsidies. His results showed that R&D subsidies greatly stimulated private R&D expenditures for small firms, but had a negative effect on large firms' R&D expenditures.

The fourth research design is propensity score matching. When using instrumental variables and selection models, the data used in the study needs to meet several strong assumptions, such as the distribution of error terms. When the data cannot meet such strong assumptions, matching is a good choice. But it only controls for observed heterogeneity between firms with subsidies and firms without subsidies. Almus and Czarnitzki (2003) used the propensity score matching approach and found that the full crowd-out effect did not exist in Eastern German manufacturing. Czarnitzki and Hussinger (2004) employed nearest neighbor matching to analyze the effect of public R&D funding on private R&D expenditures and patenting behaviors in Germany. Duguet (2004) also applied matching methods to data about France and found that the full crowd-out effect didn't exist in France. Czarnitzki et al. (2007) and González, X. and Pazó, C. (2008) also employed matching and found no crowd-out effect in Germany, Spain and Finland.

The different results about the existence of the crowd-out effect or the crowd-in effect might be due to different types of subsidies, different development stages of an industry, different subsidies amounts and various characteristics of firms. If government subsidies are at a low level, government subsidies may stimulate private investments, but when they are higher than a certain threshold, government subsidies may have the opposite effect (Gullec et al., 2000; Brooks, 2000). Tommy (2009) distinguished the "far from the market subsidies" from the "close to the market subsidies". He showed that "far from the market subsidies" can stimulate private R&D spending, but "close to the market subsidies" substitutes private R&D spending.

From the previous literature, we can see that plenty of studies investigate the relationship between public R&D investments and innovations in developed countries. But only a few studies focus on developing countries, and even fewer studies focus on emerging industries in developing countries. In order to address this gap, my study focuses on the effect of governments subsidies on firms' innovation and firms' private investments in the Chinese wind industry, and provides insights about reasons behind previous inconsistent empirical results.

Chapter 3: *Design*

To study the effect of an intervention (or a program), the best way is to compare the outcomes with and without the intervention for the same sample or population at the same time. But it is impossible to observe the counterfactual outcome, because if one unit participates in an intervention, we can never observe the outcome in the absence of the intervention for the same unit at the same time. To address this difficulty in evaluating the effect of an intervention, random experiments are very useful. We can randomly assign units into a treated group and a control group, and then directly compare the outcomes between the treated group and the control group. But the costs of designing and implementing such a random experiment are high and sometimes it is difficult to successfully accomplish random experiments.

Many programs and interventions are not assigned randomly. In order to verify the causal link between an intervention and outcomes, we need to address the problem of endogeneity and selection bias. The treatment in this study is the government subsidies, and the outcomes are firms' innovation outputs and private investments. Problems of endogeneity and selection bias both exist. In order to make full use of the subsidies and promote more innovative behaviors, governments prefer to assign subsidies to firms with high innovation capacity. On the other hand, larger and older firms, and firms with riskier innovative programs, are more willing to apply for government subsidies because they are more familiar with how to reach government subsidies. Therefore, we need to take advantage of statistical methods to remove pre-treatment.

As a nonexperimental evaluation, propensity score matching (PSM) is a popular method to evaluate the effect of a program or an intervention that is assigned non-randomly. In PSM, we can compare outcomes between participants and non-participants with similar characteristics, because I select untreated units that are similar to our treated units based on propensity scores. A very important assumption of PSM is the common support, which means untreated units that are similar to the treated units can be found in the dataset.

The PSM matches units based on propensity scores alone. Therefore, the first step is to estimate propensity scores. The dependent variable is a dummy variable indicating the treatment status (i.e., whether a unit receives a treatment or not). A logit or probit regression model is the most common method to estimate propensity scores, which is the probability that a unit receives the treatment. However, more flexible approaches, such as the tree-based approach, are considered as good alternatives to logit or probit model. Tree-based approaches are well-suited to situations in which the relationship between predictors and response variable is non-linear or more complicated. Lee et al. (2010) and Setoguchi et al. (2008) compare the logistic model with machine learning methods by using simulated data, and find that the tree-based approach performs well (or even better) in many scenarios. They also note that machine learning methods are well-suited to small sample size and high-dimensional data. The shortcoming of these flexible models is that interpreting the models can be difficult. But in propensity score matching, we care much more about matching results and treatment effects than about interpretations. The biggest challenge of the first step is to select covariates. Variables that are related to both the treatment status and outcomes should be included in the logistic regression model or the classification tree as predictor variables. In the data section, I will discuss more about how I select covariates for this study.

The second step is to decide the matching algorithm. Here we need to consider the measures of proximity and number of comparison units to each treated unit. There are three common matching algorithms. The first one is the nearest neighbor, which selects n non-participants with propensity scores closest to that of a participant. Caliper/radius matching uses all comparison units within a defined common-support region (e.g., 0.01, 0.05), not only the nearest neighbor. The third method is the kernel/local linear, which is a nonparametric matching estimator. The basic idea is to compare the outcome of each treated unit to a kernel-weighted average of the outcomes of all the untreated units (Heinrich et al., 2010). After matching, the characteristics of treated units and matched untreated units should be balanced except for treatment status, although not perfectly balanced, so we can estimate the average treatment effect by comparing the treated units with their matched untreated units. The biggest drawback of this matching method is that we cannot control for unobserved covariates. If the unobserved covariates are time-invariant, difference-in-difference models can address this problem. The other way to reduce this problem is to use pretreatment data as a covariate. In this study, I include the pretreatment data (the number of patent before 2007) as a covariate to address the problem of unobserved covariates to some extent.

The firm-level dataset comes from Chinese Annual Survey of Industrial Production (ASIP) of 2007, which is collected by China's National Bureau of Statistics. The ASIP covers the majority of industrial firms, including all state-owned enterprises and private firms with annual turnover above 5 million RMB (about U.S. $800,000). There are 170 wind firms in this dataset, which covers about 90% of all the wind firms in China in 2007. The number of patents for each firm comes from the State Intellectual Property Office of China (SIPO). I use the name of each firm as the searching keyword on the website of SIPO and collect the number of patents for each firm before 2007 and after 2007.

The two outcomes (dependent variables) in this study are the firm's private R&D investment and the number of patents after 2007. The outcome of the firm's private R&D investment aims at studying the crowd-out effect of government subsidies. The number of patents captures firms' innovation capacity. Many studies justify that the number of patents can reflect a firm's innovation capacity (Schmookler, 1962; Griliches, 1990; Lanjouw and Mody, 1996; Jaffe and Trajtenberg, 2002). The largest drawback of this measurement is that the number of patents cannot directly reflect the quality and the economic value of innovations (Griliches et al., 1987). Even though the number of patents is not a perfect proxy for a firm's innovation, it is a meaningful and valid measure.

The covariates I include in this study are related to the firms' treatment status and innovation capacity. We control for a firm's number of patents before 2007, age, location (whether the location is close to a sea), ownership (whether it is a state-owned firms or not), per capita assets (the ratio of assets to the number of employees), per capita production value (the ratio of production value to the number of employees), per new added capita production value profit in 2007 (the ratio of new added production value to the number of employees), scale, and export status (whether a firm exports their products). Table 3.1 shows summary statistics of these key variables. There are 170 firms in the dataset and 59 firms (about 35% of total firms) obtained government subsidies.

Previous researchers found that larger and older firms are more likely to obtain government subsidies (Wallsten, 2000; Almus and Czarnitzki, 2003; Czarnitzki and Hassinger, 2004; Aerts and Czarnitzki, 2004; Hassinger, 2006). But some research has found that the size of a firm is negatively related to the probability of obtaining government subsidies. Some researchers found that firm age cannot predict the probability of obtaining government subsidies. Previous studies also show that ownership can predict the probability of being subsidized. Hassinger (2006) shows affiliates of foreign firms have less probability of being subsidized. Several studies show that state-owned firms are more likely to obtain government subsidies (Hoskisson et al., 2002; Lee and O'Neill, 2003; Czarnitzki and Licht, 2006; Choi et al., 2011). Previous studies also show that exporting firms have a higher probability of obtaining subsidies (Almus and Czarnitzki, 2003; Aerts and Czarnitzki, 2004; Hassinger, 2006). The pre-treatment data measured by the number of previous patents is the most important covariate in the model. Previous studies also show the positive relationship between firms' innovation capacity and the probability of being subsidized (Wallsten, 2000; Aerts and Czarnitzki, 2004; Hassinger, 2006).

Table 3.1 Summary statistics

Variable	Obs.	Mean	Std. Dev.	Min	Max
private R&D investment	170	1177.95	10460.43	-80	134483
number of patents after 2007	170	12.52	62.68	0	563
subsidies	170	1086.43	2575.43	0	15733
subsidies status	170	0.35	0.48	0	1
number of patents before 2007	170	1.46	8.18	0	90
age	170	6.72	7.50	1	52
sea	170	0.38	0.49	0	1
ownership	170	0.08	0.27	0	1
per capita asset	170	8397.86	17666.55	0	164129.3
per capita production value	170	1859.60	3705.45	0	37506.4
per new added capita production value	170	800.26	1269.67	-1187.10	8381.20
scale	170	2.88	0.37	1	3
export	170	11237.54	83618.31	0	1046380

Chapter 4: *Methods*

To study the effect of government subsidies on firms' private R&D investment and firms' innovation, I chose to use a propensity score matching method, which is based on the study of Rosenbaum and Rubin (1983). First, I estimated the probability of being subsidized for each firm, which reduces several covariates to a one-dimensional. The dependent variable is the treatment (whether the firm obtained government subsidies). The vector of X is firms' characteristics that are related to the treatment status and outcomes. Because of the small sample size, relatively high-dimensional data, and the non-linear relationship between treatment status and outcomes, I use a classification tree to estimate the propensity score. I also use a simple additive logistic model as a comparison.

The tree-based approach uses recursive binary splitting to grow a classification tree. To be specific, a classification tree splits all the predictors at a certain cutoff that can minimize the residual sum of squares (RSS). Gareth et al. (2015)[1] describe two major characteristics of a classification tree: the first characteristic is top-down, because this method starts from the top of the tree. All the observations are in one region firstly, and then split each predictor into two branches down on the tree. The second characteristic is greedy, because when picking up a split for each predictor, we only consider this particular step without considering previous splitting steps.

The classification and regression tree (CART) is commonly implemented by the tree and rpart packages in R. The algorithm in the tree package develops from Chambers and Hastie (1991) and the algorithm in the rpart package develops from Breiman et al. (1984). The rpart package help document says that "rpart differs from the tree function mainly in its handling of surrogate variables. R package tree provides a re-implementation

[1] James, Gareth, et al. An introduction to statistical learning. Vol. 6. New York: springer, 2013.

of tree. An rpart object is a superset of a tree object" (Therneau, 2015). The results from the tree package are usually different from the results from the rpart package, because two packages use different metrics to split a tree. The tree package is based on deviance statistics. But the rpart package uses Gini coefficient (a measurement of inequality) or a measurement based on information theory[2]. The rpart package also adds a penalty of adding a new split in a tree by including a complexity parameter[3], which plays a similar role as pruning the tree.

As a comparison, this study also uses logistic regression to predict propensity scores. The estimation equation is as follows:

$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 pre_{patent} + \beta_2 age + \beta_3 sea + \beta_4 per\ capita\ asset + \beta_5 per\ capita\ pro\ value + \beta_6 scale + \beta_7 export + \beta_8 ownership + \beta_9 new\ added\ per\ capita\ pro\ value$$

Where π is the probability of obtaining government subsidies, which is also the propensity score. The pre_patent is the number of firm's patent before 2007, age is the length of years between firm's establishment and 2007, sea is a dummy variable indicating that whether a firm locates close to a sea, ownership is a dummy variable indicating that whether a firm is a state-owned firm, per capita assets is the ratio of assets to the number of employees, per capita production value is the ratio of production value to the number of employees, per new added capita production value is the ratio of new added production value in 2017 to the number of employees, export is a dummy variable indicating that whether a firm exports their products, and firm's scale has three values: scale=1 means large size, scale=2 means median size and scale =3 means small size.

---

[2] Source: Fridley Lab. Available at: http://plantecology.syr.edu/fridley/bio793/cart.html
[3] Source: Fridley Lab. Available at: http://plantecology.syr.edu/fridley/bio793/cart.html

Then I used one to one matching with replacement to match each treated firm with untreated firms, because the number of firms in the control group is only twice the number of treated firms. Then I can estimate the causal effect of government subsidies on firms' private R&D investment and firms' innovation between treated firms and their matched untreated firms, respectively.

Chapter 5: *Results*

Figure 5.1 and Figure 5.2 show results of the classification tree by using the tree and rpart package, respectively. The tree package results use 7 predictors and have 15 internal nodes. However, results by using the rpart package is a pruned tree, which uses 5 predictors and has only 8 internal nodes. Both results show that per capita assets, sea and per capita production values are the most important predictors.

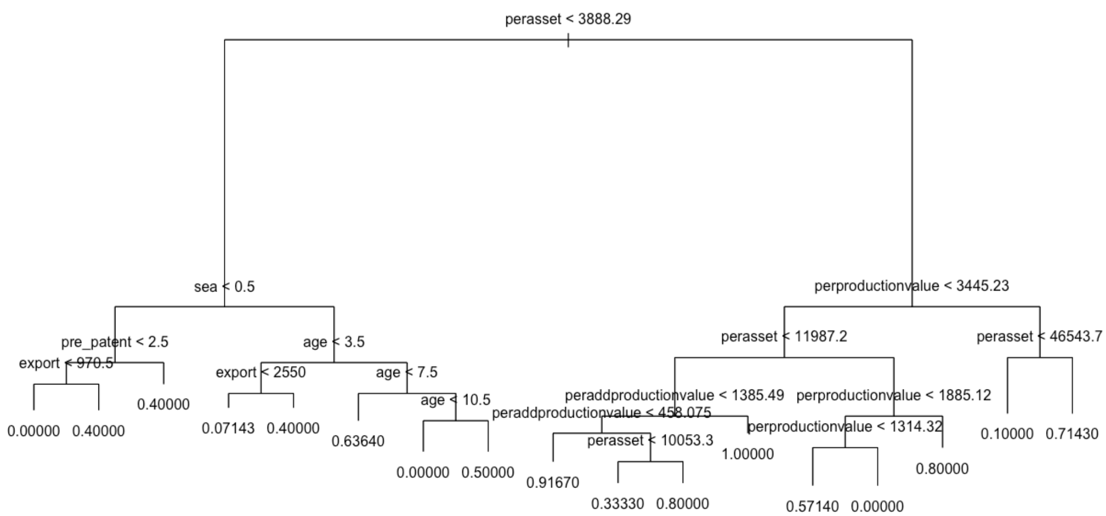

Figure 5.1 Results of the classification tree by using tree package

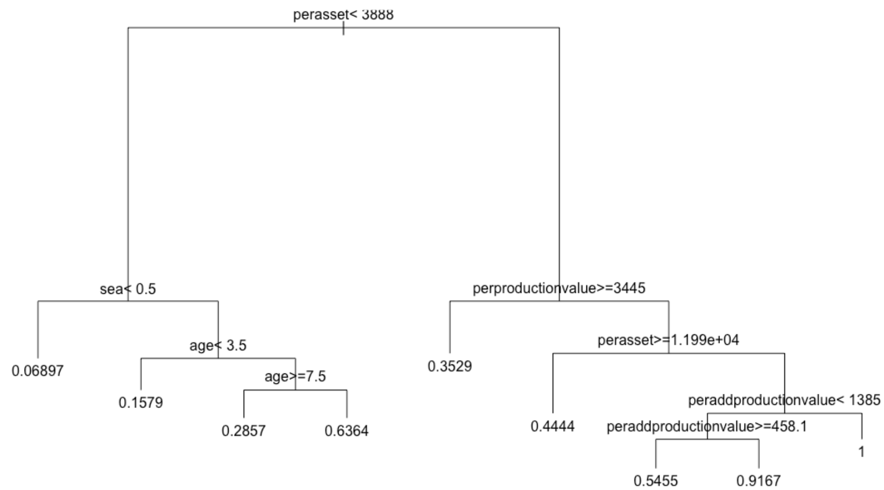

Figure 5.2 Results of the classification tree by using rpart package

Table 5.1 shows results of the logistic model. Four variables are statistically significant. The probability of being subsidized increases with *per capita* new added production value, and *per capita* asset, and decreases with *per capita* production value. Being close to the sea also increases the probability of being subsidized. Almost all the prosperous cities in China locate near the sea and many universities concentrate in these cities. Therefore, firms that locate close to the sea have advantageous conditions to develop very well. Thus, it is reasonable that these firms have a higher probability of obtaining subsidies. Regarding these statistically significant predictors, the logistic regression results are consistent with results of the classification tree. It is surprising that a firm's scale is not statistically significant, but the *per capita* scales are very important variables to predict the probability of being subsidized. It is also surprising to find that pre-treatment data (the number of patents before 2007) is not statistically significant. A possible reason is that the wind industry in China was still at the very early stages in 2007. Most firms only had a few number of patents before 2007, which cannot reflect firms' innovation capacity very well.

Table 5.1 Results of the logistic model

	Coefficient	Std. Err.	Z value	P-value
number of patent before 2007	0.01814	0.022	0.810	0.418
age	-0.01608	0.027	-0.596	0.551
sea	0.76380**	0.379	2.017	0.044
ownership	0.28110	0.674	0.417	0.676
per new added capita production value	0.00086**	0.000	1.969	0.049
per capita asset	0.00004*	0.000	1.737	0.082
per capita production value	-0.00043**	0.000	-2.354	0.019
scale	-0.80300	0.567	-1.415	0.157
export	-0.00001	0.000	-1.115	0.265
constant	1.25300	1.687	0.743	0.458
Obs	170			

*** indicates statistical significance at the 1% level; **indicates statistical significance at the 5% level; *statistical significance at the 10% level.

Based on the classification tree and logistic regression results, I can obtain a propensity score for each firm. Figures 5.3 to Figures 5.5 show the distribution of propensity scores for both treated firms and untreated firms by using the classification tree based on the tree package, the rpart package, and the logistic regression model, respectively. The results show that the more flexible classification tree provides a better opportunity for the matching, because the distribution of propensity scores is more symmetrical between treated firms and untreated firms when using the classification tree. There is almost no overlapping of propensity scores at values less than 0.2 or values around 0.5.

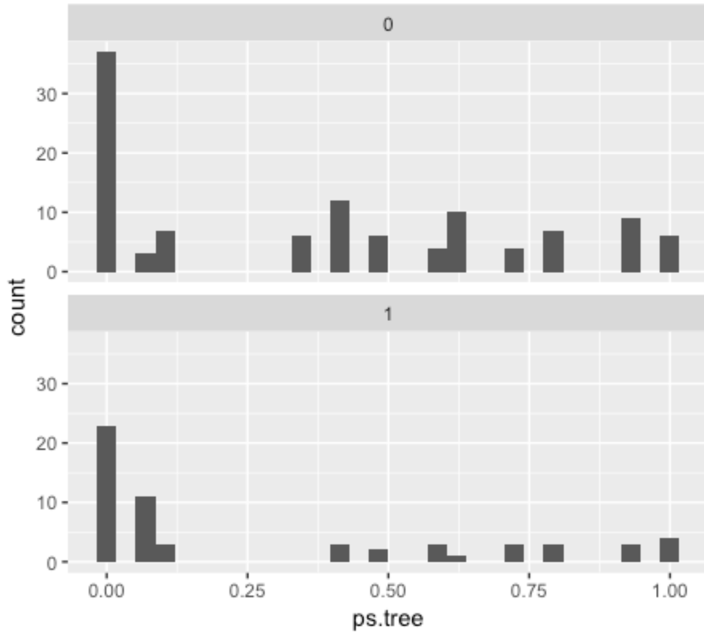

Figure 5.3 Propensity scores for treated firms (title=1) and untreated firms (title=0) based on the tree package

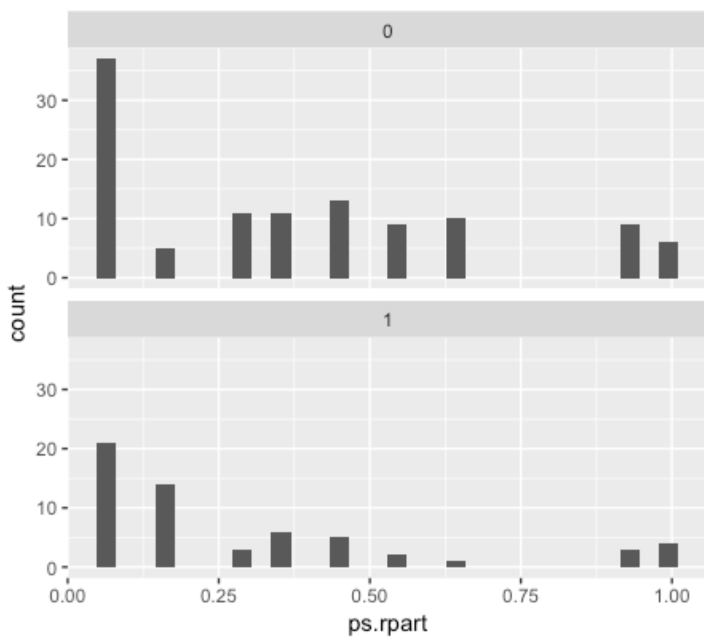

Figure 5.4 Propensity scores for treated firms (title=1) and untreated firms (title=0) based on the rpart package

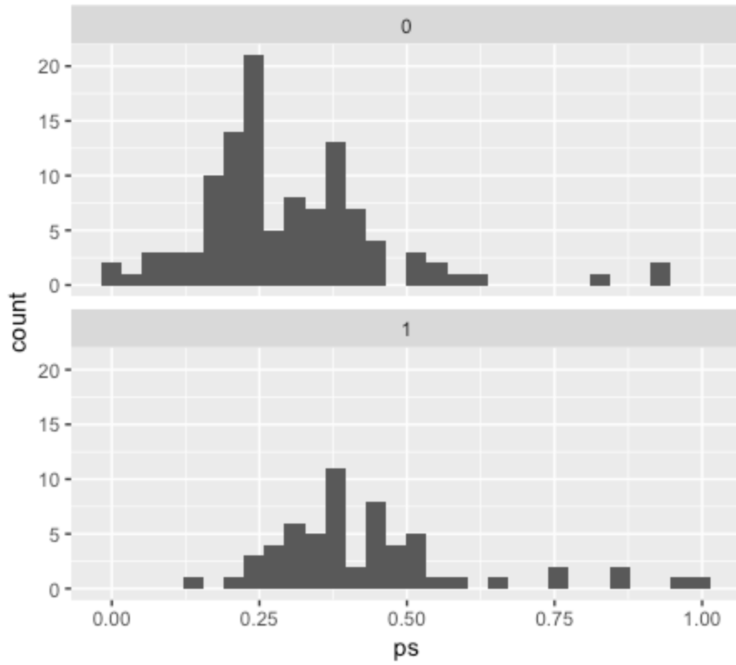

Figure 5.5 Propensity scores for treated firms (title=1) and untreated firms (title=0) based on the logistic regression results

Table 5.2 The treatment effect for the number of patents after 2007

	Logistic regression	Classification tree by using tree package	Classification tree by using rpart package
Treat mean	20.102	20.102	20.102
Control mean	12.051	11.729	2.169
Mean(D = Treat - Control)	8.051	8.373	17.932
SD(D)	106.886	100.816	75.661
Lower 95% Confidence Interval	35.906	34.646	37.65
Upper 95% Confidence Interval	-19.804	-17.9	-1.785
t (D-bar)	0.579	0.638	1.82
p-value (t-statistic)	0.565	0.526	0.074

Table 5.3 The treatment effect for private research and development investments

	Logistic regression	Classification tree by using tree package	Classification tree by using rpart package
Treat mean	2834.441	2834.441	2834.441
Control mean	185.644	116.542	88.593
Mean (D = Treat - Control)	2648.797	2717.898	2745.847
SD(D)	17624.929	17603.542	17599.024
Lower 95% Confidence Interval	7241.878	7305.406	7332.178
Upper 95% Confidence Interval	-1944.285	-1869.61	-1840.483
t (D-bar)	1.154	1.186	1.198
p-value (t-statistic)	0.253	0.24	0.236

The table 5.2 shows the average treatment effect of government subsidies on a firm's innovation capacity. I use one-to-one matching with replacement. The average treatment effect is positive, but it is not statistically significant. It means that there is no causal link between government subsidies for wind firms and wind firms' innovation outcomes. An interesting result is that results using the rpart package are very different from results using the tree package. By checking propensity scores and matching pairs based on these two methods, I find that propensity scores obtained by these two methods are different, which lead to very different matching pairs. Only 24 of 59 pairs are the same in these two methods, which might explain the different treatment effects in two methods. This interesting result is worthy of further exploration.

Table 5.3 shows the average treatment effect of government subsidies on a firm's private R&D investments. There is a positive relationship between government subsidies and a firm's private R&D investments, but it is not statistically significant as well, which means there is no crowd-out effect or crowd-in effect of government subsidies on private R&D investments in innovation activities.

These non-ideal policy outcomes may result from the immaturity of the Chinese wind industry in 2007. Since technological innovations cannot occur within a short term, maybe only long-term subsidies in larger amounts can have effects on firms' technological innovations. As for the crowd-in or the crowd-out effect, it is possible that at the early development stage of the industry, wind firms cannot predict whether the government will provide persistent subsidies for their innovation activities, so firms don't have strong incentives to invest their own money in risky and long-term innovation activities.

Chapter 6: *Conclusion*

Knowledge is a public good, which has positive externalities, so investments and innovation activities are always below the most socially efficient level. Therefore, government subsidies are necessary to stimulate firms' innovation activities. This study provides new evidence of the average treatment effect of Chinese government subsidies on Chinese wind firms in 2007 by using the propensity score matching method. I use the pre-treatment data (the number of patents before 2007) to reduce the problem of possible conditional dependence between the key predictor variable, the status of government subsidies, and outcome variables (private R&D investments and the number of a firm's patents). Based on propensity scores, common support exists between subsidized firms and non-subsidized firms. Therefore, the comparison between subsidized firms and their matched non-subsidized firms can capture the average treatment effect. The results show that the average treatment effect of government subsidies on firms' innovation capacity is not statistically significant. It also shows that there is no causal link between government subsidies and a firm's private R&D investments. A sensitivity analysis for unmeasured confounding variables is needed as a future study.

These suboptimal policy outcomes may result from the immaturity of the Chinese wind industry in 2007. Innovation is more likely to occur in the long term with a large amount of investment. Thus, the subsidies must be sufficiently large and persistent enough to stimulate innovation activities. Long-term innovation activities are always very risky, and firms will be very cautious about their decisions regarding innovation. Government subsidies can increase firms' confidence because it shows that governments strongly support the emerging clean energy industry. Therefore, governments should provide more positive signals to firms, which can increase firms' incentive to engage in innovation.

After 2008, the Chinese government increased both amounts and types of R&D subsidies for wind firms. It will be very interesting to study the causal effect of government subsidies over time and especially since the Chinese government increased subsidy amounts and kept sending positive signals to the wind industry. In this future study, propensity score matching is still a good choice for this research question. If the data about how governments assign subsidies (i.e., the threshold of scores for firms' application proposals) is available, the regression discontinuity method will be a better choice for the future study.

References

Almus, M. and Czarnitzki, D., 2003. The effects of public R&D subsidies on firms' innovation activities: the case of Eastern Germany. Journal of Business & Economic Statistics, 21(2), pp.226-236.

Borrás, S. and Edquist, C., 2013. The choice of innovation policy instruments. Technological forecasting and social change, 80(8), pp.1513-1522.

Busom*, I., 2000. An empirical evaluation of the effects of R&D subsidies. Economics of innovation and new technology, 9(2), pp.111-148.

Breiman, L., Friedman, J., Stone, C.J. and Olshen, R.A., 1984. Classification and regression trees. CRC press.

Carter, C.F. and Williams, B.R., 1957. Industry and technical progress.

Carter, C.F. and Williams, B.R., 1958. Investment in innovation.

Chambers, J. M., and Trevor J. H., 1991. Statistical models in S. CRC Press, Inc.

Choi, S.B., Lee, S.H. and Williams, C., 2011. Ownership and firm innovation in a transition economy: Evidence from China. Research Policy, 40(3), pp.441-452.

Clausen, T.H., 2009. Do subsidies have positive impacts on R&D and innovation activities at the firm level?. Structural Change and Economic Dynamics, 20(4), pp.239-253.

Czarnitzki, D. and Hussinger, K., 2004. The link between R&D subsidies, R&D spending and technological performance.

David, P.A., Hall, B.H. and Toole, A.A., 2000. Is public R&D a complement or substitute for private R&D? A review of the econometric evidence. Research policy, 29(4), pp.497-529.

Duguet, E., 2003. Are R&D subsidies a substitute or a complement to privately funded R&D? Evidence from France using propensity score methods for non-experimental data.

Fridley Lab. Available at: http://plantecology.syr.edu/fridley/bio793/cart.html

Freeman, C., 1989. Technology policy and economic performance (p. 34). Great Britain: Pinter Publishers.

Furman, J.L., Porter, M.E. and Stern, S., 2000, August. UNDERSTANDING THE DRIVERS OF NATIONAL INNOVATIVE CAPACITY. In Academy of Management Proceedings (Vol. 2000, No. 1, pp. A1-A6). Academy of Management.

González, X. and Pazó, C., 2008. Do public subsidies stimulate private R&D spending?. Research Policy, 37(3), pp.371-389.

González, X., Jaumandreu, J. and Pazó, C., 2005. Barriers to innovation and subsidy effectiveness. RAND Journal of Economics, pp.930-950.

Grossman, G.M. and Helpman, E., 1993. Innovation and growth in the global economy. MIT press.

Heijs, J. and Herrera Danny, L., 2004. The distribution of R&D subsidies and its effect on the final outcome of innovation policy.

Hottenrott, H. and Lopes-Bento, C., 2014. (International) R&D collaboration and SMEs: The effectiveness of targeted public R&D support schemes. Research Policy, 43(6), pp.1055-1066.

Huergo, E. and Moreno, L., 2014. National or international public funding? Subsidies or loans? Evaluating the innovation impact of R&D support programmes.

Hussinger, K., 2008. R&D and subsidies at the firm level: An application of parametric and semiparametric two - step selection models. Journal of applied econometrics, 23(6), pp.729-747.

Jaffe, A.B. and Trajtenberg, M., 2002. Patents, citations, and innovations: A window on the knowledge economy. MIT press.

Jaffe, A.B. and Le, T., 2015. The impact of R&D subsidy on innovation: a study of New Zealand firms (No. w21479). National Bureau of Economic Research.

James, G., Witten, D., Hastie, T. and Tibshirani, R., 2013. An introduction to statistical learning (Vol. 6). New York: springer.

Klette, T.J. and Møen, J., 2012. R&D investment responses to R&D subsidies: A theoretical analysis and a microeconometric study. World Review of Science, Technology and Sustainable Development, 9(2-4), pp.169-203.

Klette, T.J., Møen, J. and Griliches, Z., 2000. Do subsidies to commercial R&D reduce market failures? Microeconometric evaluation studies. Research Policy, 29(4), pp.471-495.

Kline, J.K. and Rosenberg, N., 1986. An Overview of Innovation in R. Landau and N. Rosenburg (eds) The Positive Sum Strategy: Harnessing Technology for Economic Growth Washington DC.

Kline, S.J., 1989. Research, invention, innovation, and production: models and reality. Thermosciences Division, Department of Mechanical Engineering, Stanford University.

Lach, S., 2002. Do R&D subsidies stimulate or displace private R&D? Evidence from Israel. The journal of industrial economics, 50(4), pp.369-390.

Lee, B.K., Lessler, J. and Stuart, E.A., 2010. Improving propensity score weighting using machine learning. Statistics in medicine, 29(3), pp.337-346.

Lee, E., Walker, M. and Zeng, C., 2014. Do Chinese government subsidies affect firm value? Accounting, Organizations and Society, 39(3), pp.149-169.

Raiteri, E. and Guerzoni, M., Innovative Procurement and R&D Subsidies: confounding effect and new empirical evidence on technological policies in a quasi-experimental setting.

Rosenbaum, P.R. and Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika, pp.41-55.

Rothwell, R., 1983. Innovation and firm size-a case for dynamic complementarity-or, is small really so beautiful. Journal of General Management, 8(3), pp.5-25.

Rothwell, R., 1989. Small firms, innovation and industrial change. Small Business Economics, 1(1), pp.51-64.

Setoguchi, S., Schneeweiss, S., Brookhart, M.A., Glynn, R.J. and Cook, E.F., 2008. Evaluating uses of data mining techniques in propensity score estimation: a simulation study. Pharmacoepidemiology and drug safety, 17(6), pp.546-555.

Wallsten, S.J., 2000. The effects of government-industry R&D programs on private R&D: the case of the Small Business Innovation Research program. The RAND Journal of Economics, pp.82-100.

Wales, W.J., Parida, V. and Patel, P.C., 2013. Too much of a good thing? Absorptive capacity, firm performance, and the moderating role of entrepreneurial orientation. Strategic Management Journal, 34(5), pp.622-633.