# Effective Only to Low Capital Firms – Regression Discontinuity Replication Analysis on the Bronzini and Iachini Paper on R&D Subsidy Program in Northern Italy

Yixin Guan

April 24 2021

**Abstract**

This paper re-evaluates Bronzini and Iachini's paper on the R&D subsidy program implemented in northern Italy using sharp regression discountinuity design. Bronzini and Iachini compared the investment spending of subsidized firms with that of unsubsidized firms and did not find a significant increase in investment. This paper seperates the firms into two groups based on their pre-program total capital status examine the effectiveness of the program in supporting firms' research and development investment. This paper found that the program is effective to firms with low pre-program capital reserves but not effective to firms with high pre-program capital.

**Keywords:** Regression Discontinuity Design, Bandwidth, Governmental subsidy, Innovation policy, Financial support

## 1 Introduction

The use of public direct funding as incentives to boost industry research and development (R&D) is an important component of the economic development agenda for most governments around the world. The amount of financial resources involved can be substantial and is subject to public scrutiny. The effectiveness of the programs is a suggestion of public accountability and public trust.

Notwithstanding the popularity of R&D investment subsidies, the question of whether they actually work - i.e., increase firms' R&D activity - remains unsettled. Bronzini and Iachini's paper (Bronzini and Iachini 2014) studied a unique program implemented in a region of northern Italy (Emilia-Romagna). The policy has several key features enabling the effectiveness of R&D incentives to be carefully assessed. First, it allows them to address the endogeneity issue with sharp regression discontinuity (RD) design. The program envisages that only eligible projects that receive a higher than 75 out of 100 points in an assessment by an independent technical committee will be subsidized. Bronzini and Iachini's paper compare the investment of subsidized and unsubsidized firms close to the threshold score. Secondly, the program's local dimension allowed them to remove the unobserved heterogeneity among firms to a certain extent and compare recipients and nonrecipients that are more similar. A third advantage is the program's size (overall, about €93 million have been granted) and its involvement of a large number of firms (1 ,246 applicant enterprises). In their baseline sample, each subsidized firm received an average of €1 82,000 - one-fourth of the total investment made by each participating firm in the two years after the program. The size of the grants and the high participation rate are helpful for the evaluation exercise. The Bronzini and Iachini's paper took the assessment score from the independent technical committee as the independent variables (assuming the decision of the committee as made independent of the company). From the balance sheets Bronzini and Iachini take as outcome variables those items that are associated with the expenditures reimbursable by the program. The rationale is that if the program enabled outlays that without the grant would not have been made, they should observe a significant increase in at least one of program funding areas for the recipient firms after the program, compared with those of nonrecipient firms.

However, the Bronzini and Iachini's paper suffers some limitations. The Bronzini and Iachini's paper classified the firms into small firms (n=178) and large firm (n=179) and found that subsidies are effective to the small firms but not the big firms. However, their paper but did not explain the criteria used for the classification. The subsidy program provides grant that may cover up to 50 percent of the costs of industrial research projects and 25 percent for precompetitive development projects, with the remain funding provided by the firm themselves. Whether the firms are have enough capital to make up the remaining funding needed may impact their ability to take advantage of the funding program and may skew the results. Therefore, this paper will re-examine the analysis of this paper by separating firms into two groups based on their pre-program total capital. The firms with pre-program total capital below the median contains 26 large firms and 153 small firms, and firms with pre-program total capital above the median contains 152 large firms and 26 small firms. This paper uses the same methodology from Bronzini and Iachini's paper to examine the effect of the subsidy to firms based on their pre-program total capital categories and their investment situation after receiving the subsidy from the program.

This paper is structured as follows. In Section I, I illustrate the features of the program. In section II, I discuss the data used in this paper. Section III provides a introduction of the model used in this paper, followed by the presentation of analysis results in Section IV. Section V will discuss the results and the limitation of data and the analysis methods. This paper will end with a conclusion in Section VI.

The Github repo can be found in the footer.[1]

# 2 Data

This paper uses the same dataset provided by the original Bronzini and Iachini's paper. Bronzini and Iachini's paper conducted their analysis using STATA (StataCorp 2003), while this paper conducts the analysis with R (R Core Team 2020). R packages including 'haven' (Wickham and Miller 2021), `tidyverse` (Wickham et al. 2019), `here` (Müller 2020), `kableExtra` (Zhu 2020),`modelsummary` (Arel-Bundock 2021), `gplot2` (Wickham 2016),`gcookbook` (Chang 2018) were used to conduct the analysis and present results. Due to the use of different statistical tools, the some results presented in this paper may slightly vary from the original paper. This is considered as acceptable.

## 2.1 The program

The paper studies the "Regional Program for Industrial Research, Innovation and Technological Transfer" by the government Emilia-Romagna, Italy in 2005 (Muscio 2014). As indicated in the Bronzini and Iachini's paper, through this program, regional government provides direct R&D subsidy to eligible firms to support their industrial research and precompetitive development. The program grant may cover up to 50 percent of the costs of industrial research projects and 25 percent for precompetitive development projects; the 25 percent limit is extended by an additional 10 percent if applicants are small or medium-sized enterprises in the region. Several types of outlay related to the eligible project can be subsidized: (i) costs for machinery and equipment; (ii) software; (iii) purchase and registration of patents and licenses; (iv) employment of researchers; (v) the use of laboratories; (vi) contracts with research centers; (vii) consulting; (viii) feasibility studies; and (ix) external costs for the realization of prototypes. The maximum grant per project is €250, 000. The industrial firms data in Bronzini and Iachini's sample, grants averaged €182,000. The duration of the investment is from 12 to 24 months, but it can be extended. Subsidies are transferred to the firms either after the completion of the project, or in two installments: the first at the completion of 50 percent of the project and the second on completion. Firms that received the program grant cannot receive other types of public subsidy for the same project (Bronzini and Iachini 2014).

The assessment of the grant eligibility is conducted by a panel of independent experts appointed by the regional government. The panel assigns a score for each of the following aspects: (i) technological and scientific (max. 45 points) ; (ii) financial and economic (max. 20 points); (iii) managerial (max. 20 points); and (iv) regional impact (max. 15 points).Only projects deemed sufficient in each category and which obtain a total score of at least 75 points receive the grants (the maximum score is 100). This assessment process indicates that the applicants do not have the power manipulate their score.

[1] https://github.com/Yixin1103/final_assignment_project

## 2.2 Variables

Same as the Bronzini and Iachini's paper, this paper examine the effectiveness of the Program in increasing firms' R&D outlays by analyzing whether the subsidized firms would have made the same amount of R&D outlays without the grant. The paper uses the regression discontinuity design (RDD) method to conduct the analysis, which would be introduced in detail in the next section.

The independent variable used in the model is the assessment score given be the independent committee to each companies. The variable is continuous integer ranging from 0 to 100. Firms with scores of 75 and above are subsidized and those with scores less than 75 are not subsidized. As shown in the density distribution score in (Figure 1), the sample contains more firms that are subsidized than those that are not. The dataset includes 357 firms, of which 254 firms are subsidized, and 103 are not subsidized. What's more, 75-90 are roughly the most frequent scores companies receive.

Three outcome variables are used for the analysis, namely total investment/pre-program sales, tangible investment/pre-program sales, and intangible investment/pre-program sales. In the ideal state, the dependent variable for this analysis should be the firms R&D outlays. However, such data is not available because the data for firms' spending in R&D are not documented. Bronzini and Iachini's paper instead used the balance-sheet data that are associated with the expenditures reimbursable by the program listed in "The Program" section for the second year after the program date. The data is provided by Cerved group, which collects information on almost all Italian corporations.

Tangible investment are inputted as outcome variable because in Italy a large share of firms' expenditure for innovation activity (about 40 percent) is covered by tangible investment, such as costs for machinery and equipment (Statistica 2010). As for intangible investment, Bronzini and Iachini recognize that only part of the intangible investment are related to a firms R&D activities and items such as goodwill may be irrelevant. However, a breakdown of the intangible investment is not available for all firms from the Cerved datasets. In order to examine this issue, Bronzini and Iachini examined five participants with the largest intangible investment and found that patents, software and other intellectual property rights, license, trademark, and ongoing intangible investment cover on average about 66 percent of the total intangible investment. Therefore, they assumes that "the vast majority of the intangible investment are expenditures eligible for the subsidy" and the use of intangible investment as the dependent variable is reasonable (Bronzini and Iachini 2014). The paper employs net investment calculated from the balance-sheet data as annual differences in tangible or intangible investment net of amortization. To adjust the variances caused by the different sizes of the company, the tangible and intangible investment are adjusted based on the firm's pre-program sales. These variables are accumulated from the year of the assignment up to two years afterwards. The third outcome variable is the total investment (sum of tangible and intangible investment)/pre-program sales. This is used to capture the nuance of firms tangible and intangible investment.

This paper recognizes that labour costs, employment, wages and service costs may also be reimbursable by the Program. But these variables are not included in the analysis, since they are less direct in reflecting R&D activities of firms and it is less likely to distinguish firms' R&D spending from other spending. This paper will discuss the limitations of the variable selection in the Discussion section.

The independent and dependent variables are all consecutive ratio variables. Scatter plots of the firms' score and the three dependent variables are shown in the Appendix.

## 2.3 Capital-sufficient firms

In the original paper, Bronzini and Iachini found no significant impact of program to firms' R&D investment when examining the firms all together. However, when dividing the firms into two groups by the value added of the firms (revenue - cost), Bronzini and Iachini found that the program is effective for small firms but not the big firms.

This paper aims to contribute to the existing research by examine whether the program has different effect to firms with different capital status. The program provides 50% of the costs of industrial research projects and 25 percent for precompetitive development projects. This means that firms receiving the subsidy from the government need to match the remaining funding from their own financial accounts. According to analysis of the program setup and existing literature of Japaneses firms (David, O'Brien, and Yoshikawa 2008), this paper wants to examine whether firms with different capital structure may benefit differently from the program, i.e. show different patterns in R&D outlays. Therefore, the firms are divided into two groups, the High Capital group, whose pre-program total capital (one year before the program) is above the median, and the Low Capital group, whose pre-program capital is below the median.

(Figure 2), (Figure 3) and (Figure 4) scatter plots show the relationship between high capital and low capital firms' scores and the three outcome variables, namely total investment/pre-program sales, tangible investment/pre-program sales, and intangible investment/pre-program sales. The trend is consistent with the three outcome variables: for high capital group, firms with higher score tend to have higher R&D outlays, and for low capital group, firms with higher scores tend to have lower R&D outlays. The statistical significance these relationships requires further examination. We also observes that the high capital groups and low capital groups are somewhat evenly for treatment and comparison groups. Take the total investment group as an example, XXXXXXXXXXX

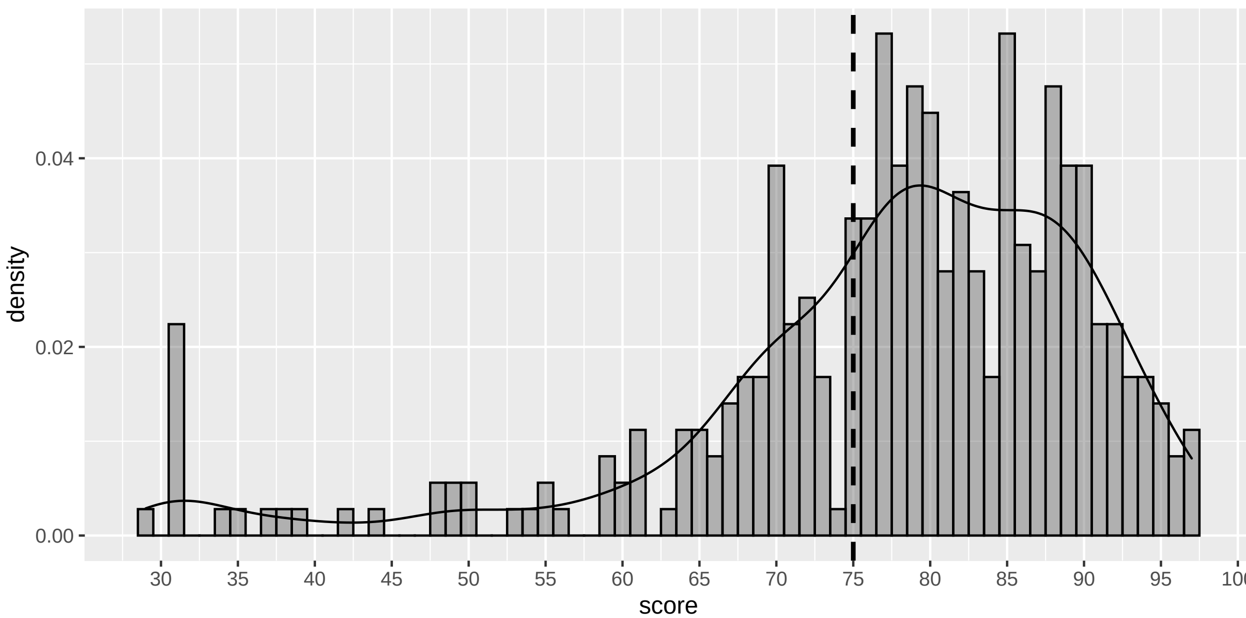



Figure 1: Firm Density Distrribution by Score

## 2.4 Bandwidth and Cutoff

Since the model in this paper is RDD, the cutoff and bandwidth needs to be defined. Firms with a score of 75 and above will be subsidized, making score 75 the sharp cutoff.

To analyze the effectiveness of the program in supporting R&D, the treatment and comparison groups should be selected to maintain the similarity between the two groups (Gertler et al. 2016). Since this dataset contains only 357 observations and was skewed towards the right of the cutoff, three bandwidth values were selected to maintain statistical power of the model, namely, the full sample, wide-window bandwidth (which contains 50% of the sample in each side of the cutoff) and the narrow-window bandwidth (which contains 35% of the sample in each side of the cutoff).

This paper tests difference between the control and treatment groups for both the low and high capital firms by examining their pre-program assets, pre-program capital and pre-program sales. As shown in (Table **??**) low capital firms, the control and treatment groups do no show statistically mean difference with the narrow window bandwidth (score 62-79), which suggest that the control and treatment groups for narrow window bandwidth do not differ significantly in pre-program assets, pre-program capital and pre-program sales. The

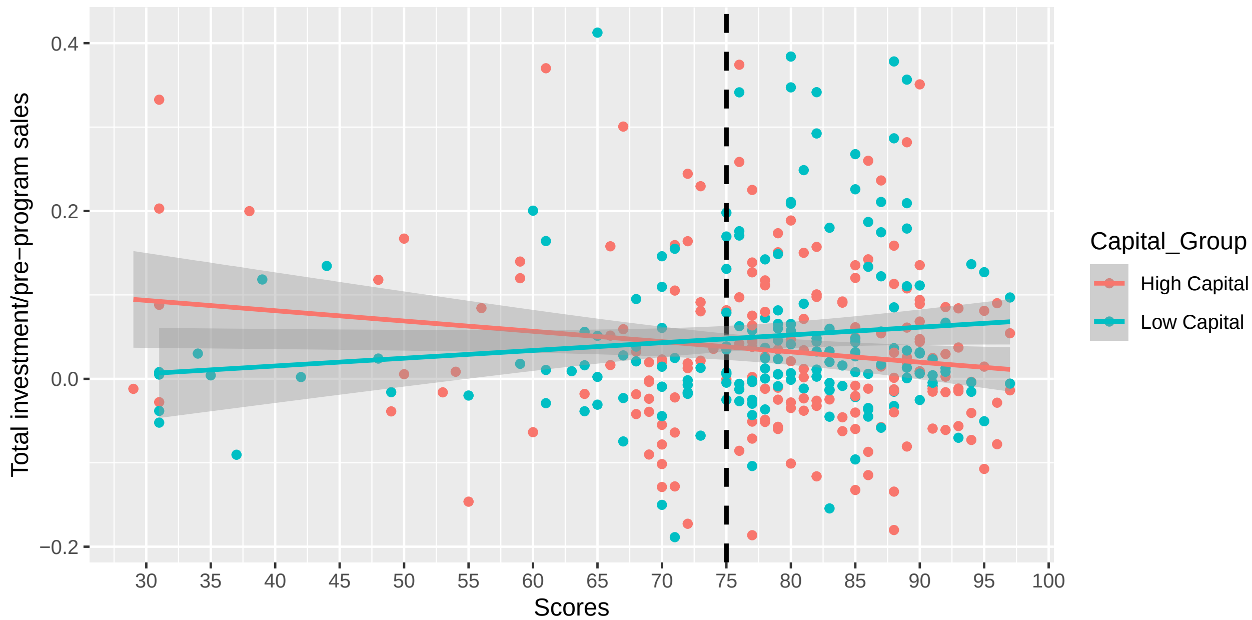



Figure 2: Score and Total investment/pre-program sales distribution

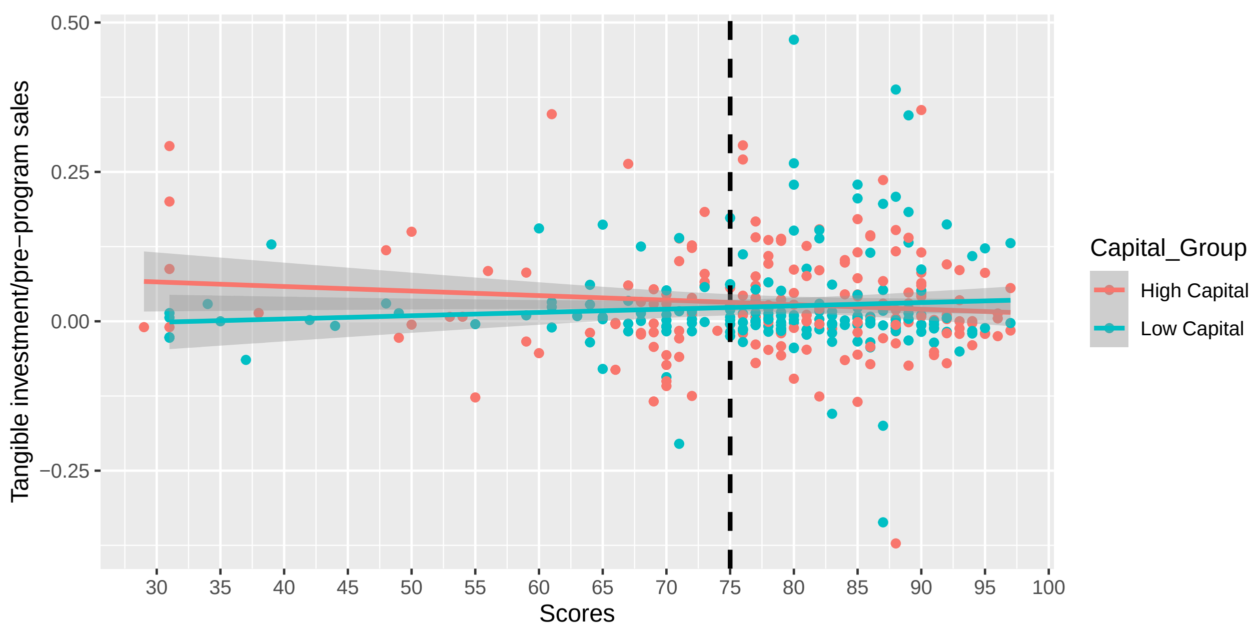



Figure 3: Score and Tangible investment/pre-program sales distribution

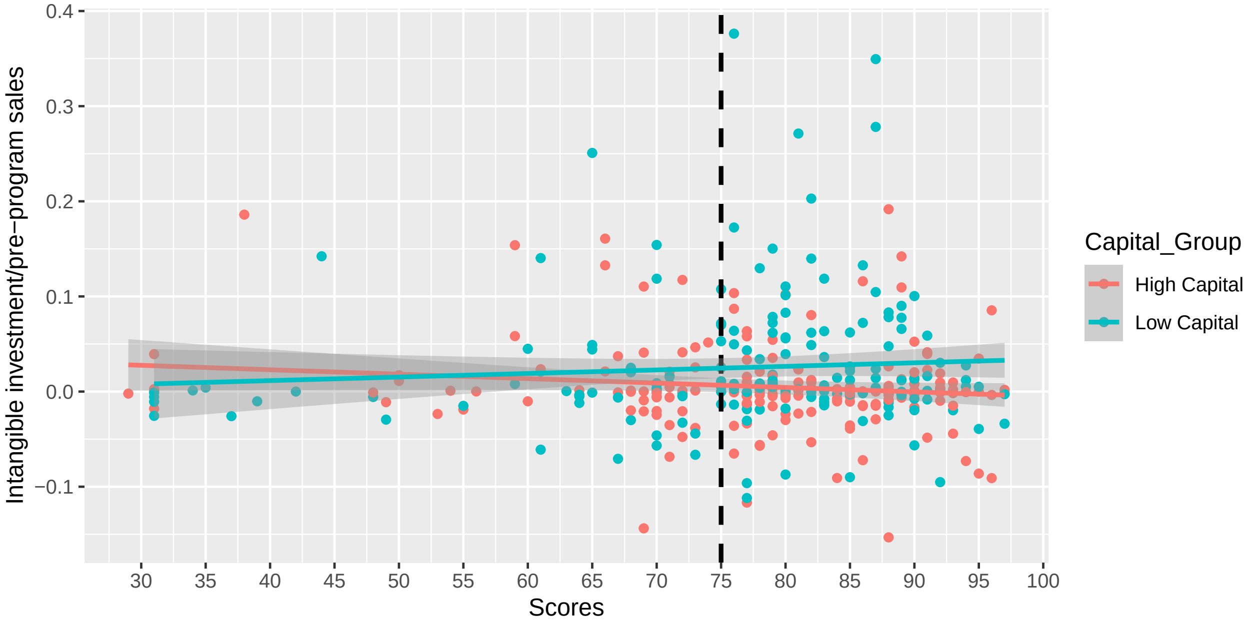



Figure 4: Score and Intangible investment/pre-program sales distribution

paper considers the two groups similar in nature. However, the treatment and control groups for the full window showed significant difference in sales and capital, and the treatment and control groups for the wide window (score 34-81) showed significant difference in sales and assets. Therefore, comparing these groups may undermine the internal validity of the RDD model.

As shown in (Table **??**) low capital firms, the control and treatment groups do no show statistically mean difference with the narrow window bandwidth (score 61-79), and the wide window (score 52 -80), which suggest that the control and treatment groups do not differ significantly in pre-program assets, pre-program capital and pre-program sales. The paper considers the two groups similar in nature. However, the treatment and control groups for the full window showed significant difference in sales, assets and capital. Therefore, comparing these groups may undermine the internal validity of the RDD model.

Based on the analysis above, the paper will analyze the high capital and low capital firms in their narrow window to examine the effectiveness of the program.

Table 1: Pre-Assignment Mean Difference between Untreated and Treated Low Capital Firms

| Variabel | Mean-Difference | statistic | P-value | 95% CI Low | 95% CI High | Method |
|---|---|---|---|---|---|---|
| **Full Window** | | | | | | |
| Sales | -3634.1785 | -3.748961 | 0.0002407 | -5547.2894 | -1721.067501 | Welch Two Sample t-test |
| Capital | -158.2338 | -2.398275 | 0.0184765 | -289.2597 | -27.207978 | Welch Two Sample t-test |
| Assets | -2755.8035 | -2.807125 | 0.0056115 | -4694.3929 | -817.214106 | Welch Two Sample t-test |
| **Wide Window** | | | | | | |
| Sales | -5336.3527 | -2.753006 | 0.0078863 | -9217.0421 | -1455.663373 | Welch Two Sample t-test |
| Capital | -113.3127 | -1.297342 | 0.1977589 | -286.7830 | 60.157553 | Welch Two Sample t-test |
| Assets | -3428.9991 | -2.062965 | 0.0429722 | -6746.3527 | -111.645513 | Welch Two Sample t-test |
| **Narrow Window** | | | | | | |
| Sales | -5360.3226 | -1.995362 | 0.0539361 | -10816.5088 | 95.863654 | Welch Two Sample t-test |
| Capital | -223.6452 | -1.988704 | 0.0514837 | -448.7789 | 1.488544 | Welch Two Sample t-test |
| Assets | -2395.0968 | -1.426707 | 0.1597099 | -5764.6961 | 974.502558 | Welch Two Sample t-test |

Table 2: Pre-Assignment Mean Difference between Untreated and Treated High Capital Firms

| name | Mean-Difference | statistic | P-value | 95% CI Low | 95% CI High | Method |
|---|---|---|---|---|---|---|
| **Full Window** | | | | | | |
| Sales | -89232.171 | -3.4402602 | 0.0007685 | -140517.29 | -37947.050 | Welch Two Sample t-test |
| K | -21657.118 | -3.1998070 | 0.0017180 | -35043.89 | -8270.343 | Welch Two Sample t-test |
| ASSETS | -67686.862 | -2.8754743 | 0.0045705 | -114166.52 | -21207.207 | Welch Two Sample t-test |
| **Wide Window** | | | | | | |
| Sales | -13687.689 | -0.8414260 | 0.4042255 | -46383.72 | 19008.345 | Welch Two Sample t-test |
| K | -3141.999 | -0.6580276 | 0.5142125 | -12786.18 | 6502.186 | Welch Two Sample t-test |
| ASSETS | -5580.259 | -0.2911419 | 0.7719387 | -43913.13 | 32752.610 | Welch Two Sample t-test |
| **Narrow Window** | | | | | | |
| Sales | -22005.190 | -1.1252069 | 0.2683434 | -61741.69 | 17731.309 | Welch Two Sample t-test |
| K | -5716.855 | -0.9347077 | 0.3575070 | -18215.22 | 6781.510 | Welch Two Sample t-test |
| ASSETS | -12254.483 | -0.5418979 | 0.5906522 | -57843.95 | 33334.988 | Welch Two Sample t-test |

# 3 Model

## 3.1 Regression Discoutinuity Design

To test the effectiveness of the Program in boosting firms' R&D investment, the most ideal way is to randomly select two groups of companies with the same qualities and provide funding to one of the group. However, this can be difficult to perform in real life. This paper, same as the Bronzini and Iachini paper, instead chooses the quasi-experiment RDD method to examine the program impact given that the firms are selected for treatment of receiving subsidies based on whether their score falls above or below 75.

The model can be used in this scenario as it satisfies the assumptions of RDD raised by Gertler (Gertler et al. 2016). Firstly, the scores of the firms are continuous and the cutoff (score = 75) is clearly defined. Secondly, firms cannot freely manipulate their score since the score is given by an independent panel of experts. We have found that firms with score of 74 are apparently low in density, but this is not considered to invalidate the design. The drop in density of firms with score 74 is likely due to the experts avoiding appeals against the decisions,and this is less likely caused by firms manipulating the score. Lastly, section 2.4 has shown that the treatment and comparison units selected are of sufficient similarities for the RDD method. This paper do recognize that due to the limitation in sample size and the availability of the outcome variable, the model may face some limitations. This will be discussed in detail in Discussion section. Overall, this paper believes that RDD is the ideal model to conduct this research.

## 3.2 Model Construction

Taking all of the above into consideration, the paper proposes the following model to evaluate the impact the program has on firms' R&D investment two years from the program start date:

$$Y_i = \alpha + \beta T_i + (1 - T_i) f(S_i) + T_i f'(S_i) + \epsilon_i$$

where

- $Y_i$ is the R&D investment outcome for firm $i$, which is represented by variables (1) Total investment/pre-program sales, (2) Tangible investment/pre-program sales, (3) Intangible investment/pre-program sales.
- $T_i$ represents whether firm $i$ has been subsidized or not. T_{i} = 1 if firm $i$ is subsidized (score_{i} > = 75) and T_{i} = 0 if firm $i$ is not subsidized (score_{i} t < 75). And its coefficient $\beta_0$ is the estimate of the program's impact on firms R&D investment when they receive the subsidies.

- $S_i$ is Score_{i} - 75, and the functions $f(S_i)$ and $f'(S_i)$ are allowed to be different on the opposite side of the cutoff to allow for heterogeneity of the function in the left and right side of the cutoff.
- $\epsilon_i$ is the random error of the function.

Bronzini and Iachini estimate up to a third-order polynomial model in their paper and the Akaike Information Criterion (AIC) value for the models suggested a preference for more simple models. Since the method used in this paper is the same as that in the Bronzini and Iachini's paper, the paper uses the first order polynomial for the model construct.

As discussed in the Data section, these outcome variables are continuous, therefore, this paper will use the linear regression model for the analysis. This paper will use the model to examine the high capital group and the low capital group sepeately.

# 4 Results

## 4.1 Low Capital Group

For firms with low capital, the coefficients for the treatment variable T_{i} are shown in the "subsidized" row in (Table **??**). The $\beta_0$ coefficient for T_{i} is significant in the total investment/pre-program sales model and the intangible investment/pre-program sales model but not on the tangible investment/pre-program sales model at a 95% confidence interval. Subsidizing a firm through the program will lead to a 0.101 increase in the firm's total investment/pre-program sales and a increase of 0.068 in the firm's intangible investment/pre-program sales. (Figure **??**) shows the program's influence to firms in the low capital groups graphically.

Table 3: Effect of the Program on Investment–Low Capital Firms

| | Total Investment | Tangible Investment | Intangible Investment |
|---|---|---|---|
| Constant | -0.028 | -0.009 | -0.020 |
| | 0.041 (0.493) | 0.024 (0.717) | 0.032 (0.543) |
| Founded | 0.101** | 0.033 | 0.068* |
| | 0.049 (0.044) | 0.029 (0.249) | 0.038 (0.081) |
| Num.Obs. | 62 | 62 | 62 |
| R2 | 0.073 | 0.024 | 0.056 |
| R2 Adj. | 0.025 | -0.027 | 0.007 |
| AIC | -108.2 | -175.4 | -139.2 |
| Log.Lik. | 59.122 | 92.686 | 74.590 |
| F | 1.512 | 0.472 | 1.138 |

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

[1] Odds Ratio in First Row. Standard Error of Log of Odds in Second Row. P-value in Parentheses

[2] Addjusted for pre-program sales. Therefore, the variables are Total Investment/pre-program sales, Tangible Investment/pre-program sales, Intangible Investment/pre-program sales

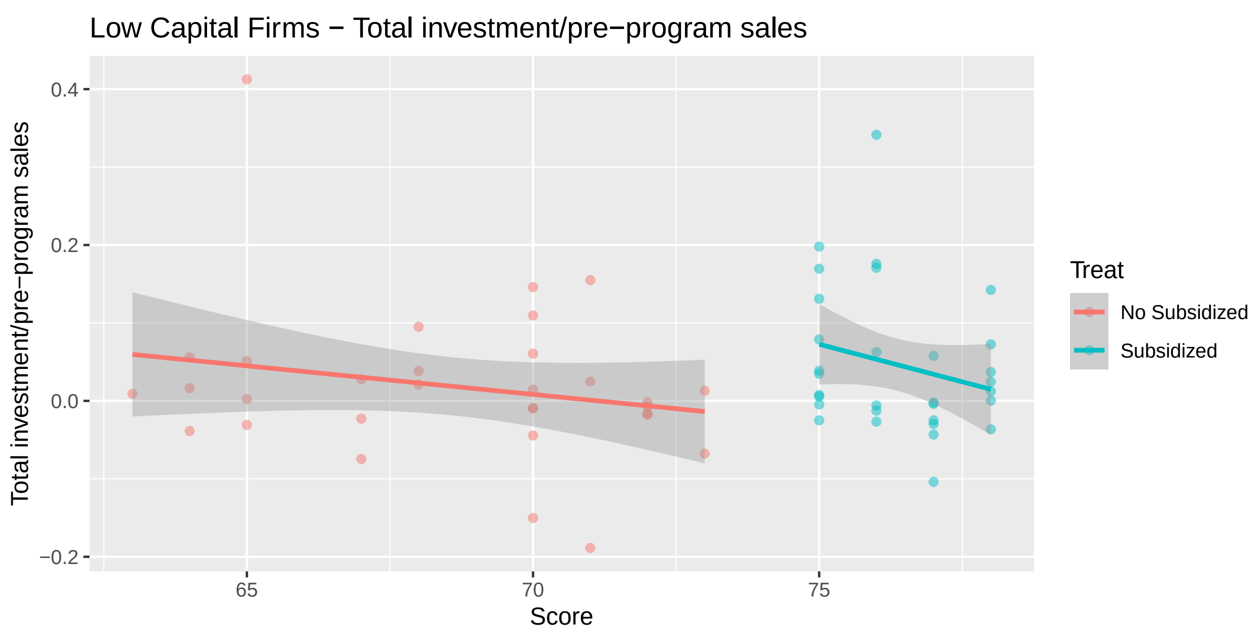
Table 4: Effect of the Program on Investment–High Capital Firms

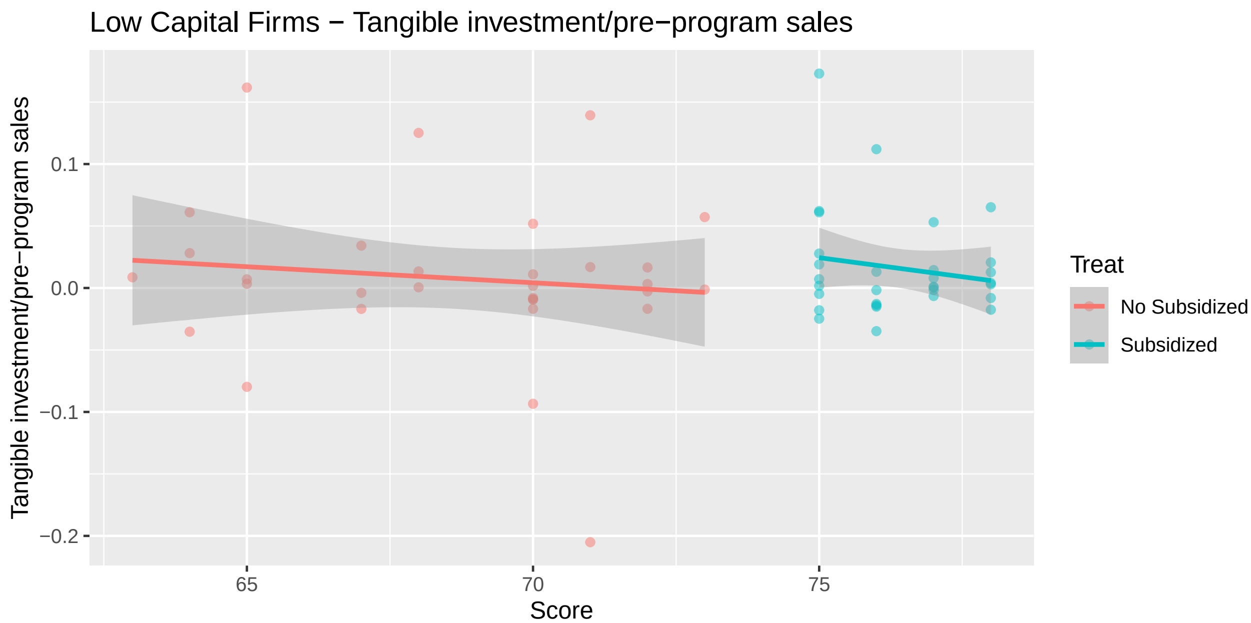
| | Total Investment | Tangible Investment | Intangible Investment |
|---|---|---|---|
| Constant | 0.035 | 0.048 | -0.013 |
| | 0.045 (0.436) | 0.036 (0.186) | 0.021 (0.545) |
| Founded | 0.069 | 0.031 | 0.038 |
| | 0.069 (0.326) | 0.055 (0.579) | 0.033 (0.250) |
| Num.Obs. | 62 | 62 | 62 |
| R2 | 0.037 | 0.071 | 0.050 |
| R2 Adj. | -0.012 | 0.023 | 0.000 |
| AIC | -90.4 | -118.4 | -184.2 |
| Log.Lik. | 50.190 | 64.182 | 97.080 |
| F | 0.750 | 1.480 | 1.010 |

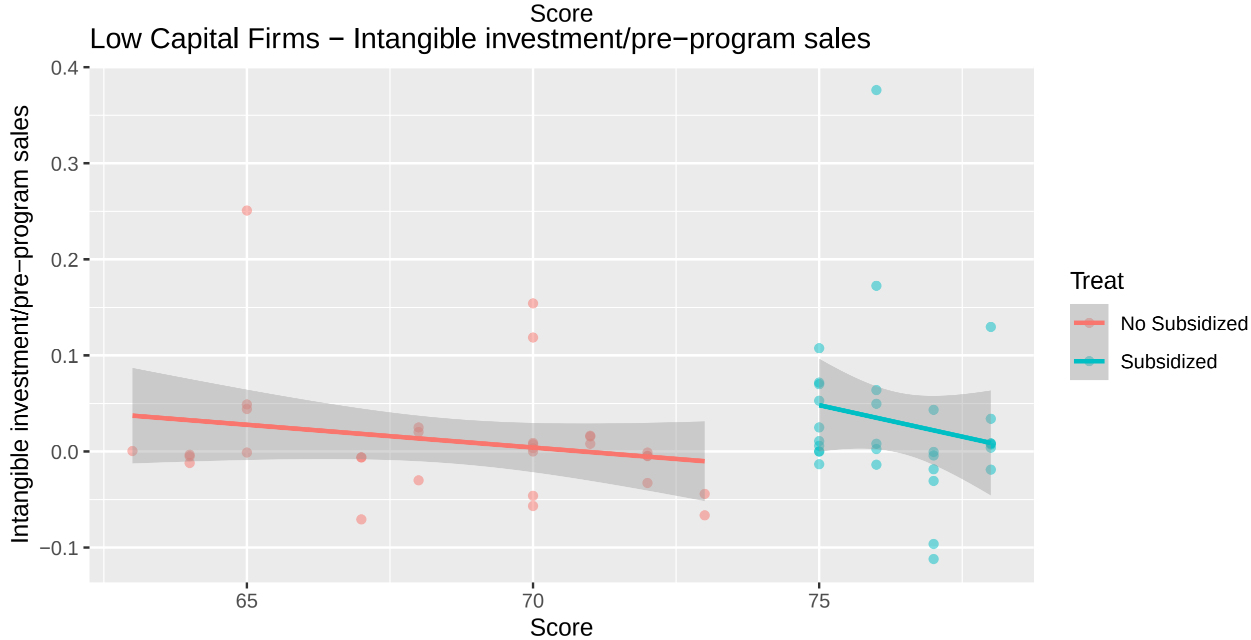
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

[1] Odds Ratio in First Row. Standard Error of Log of Odds in Second Row. P-value in Parentheses

[2] Addjusted for pre-program sales. Therefore, the variables are Total Investment/pre-program sales, Tangible Investment/pre-program sales, Intangible Investment/pre-program sales





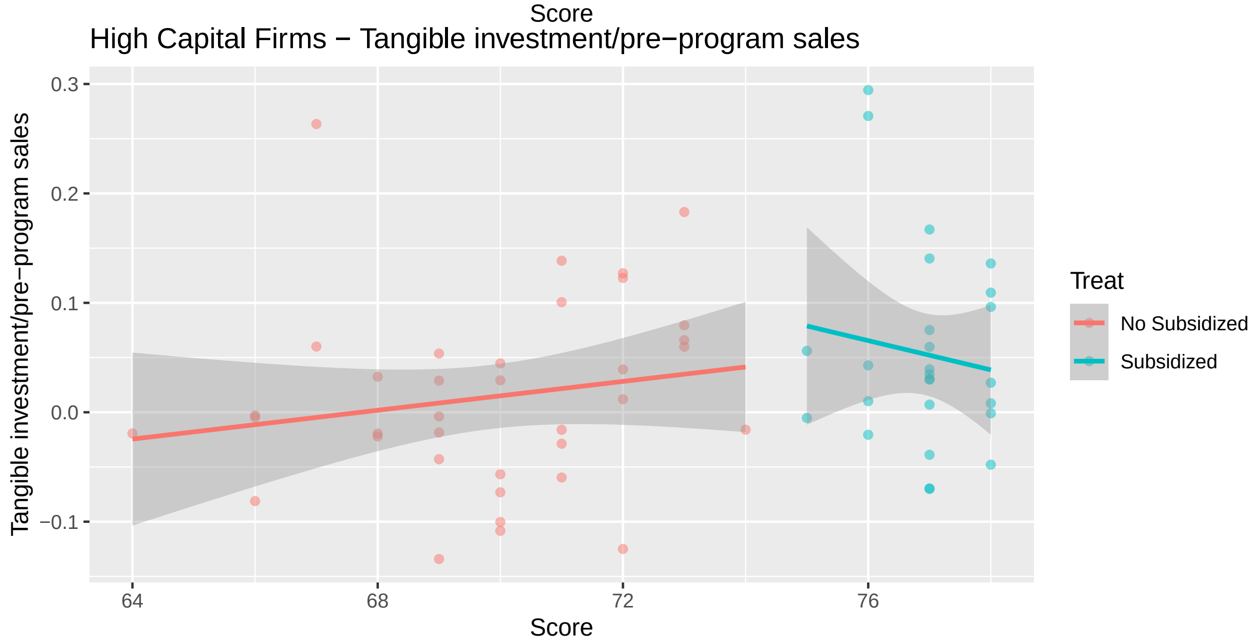



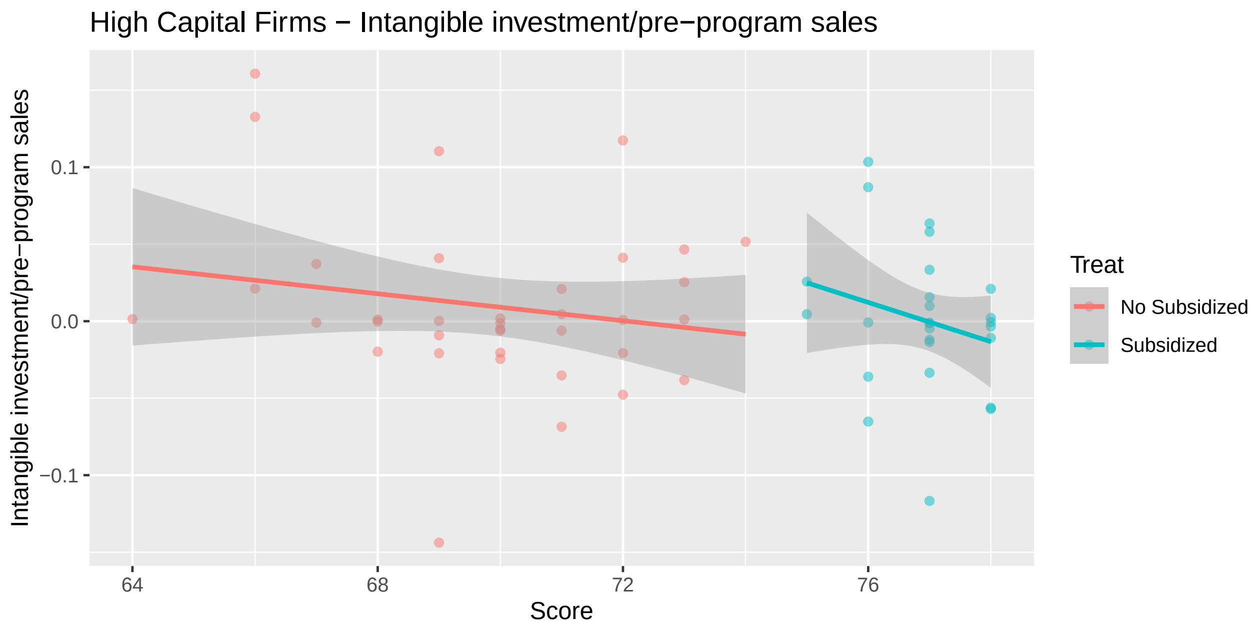

## 4.2 High Capital Group

For firms with high capital, the coefficients for the treatment variable T_{i} are shown in the "subsidized" row in (Table **??**). The $\beta_0$ coefficient for T_{i} is not significant in in neither the total investment/pre-program sales model, the intangible investment/pre-program sales model nor the tangible investment/pre-program sales model. This means that subsidizing a firm through the program will not lead to statisitically significant change in the firm's total investment/pre-program sales, intangible investment/pre-program sales, or tangible investment/pre-program sales. (Figure **??**) shows the program's influence to firms in the high capital groups graphically.

Readers can also interact with the graphical data via the shinyapps at https://yixin-guan.shinyapps.io/companygraph/.


High Capital Firms − Total investment/pre−program sales
0.4
0.2
0.0
−0.2
Total investment/pre−program sales
64
68
72
76
Score
Treat
No Subsidized
Subsidized


High Capital Firms − Tangible investment/pre−program sales
0.3
0.2
0.1
0.0
−0.1
Tangible investment/pre−program sales
64
68
72
76
Score
Treat
No Subsidized
Subsidized




The interpolations line is almost flat, the indicates a weak dependence of the overall outcome on the score. There has not been a remarkable increase of the outcome variables at the 75 cutoff points shown in the figure. We found that

# 5 Discussion

## 5.1 Key findings

This paper aims to reproduce the analysis done by Bronzini and Iachini to examine the impact of the government of Emilia-Romagna's "Regional Program for Industrial Research, Innovation and Technological Transfer" on firms' R&D investment in the next two years. Given that the program only provides funding to 50% of the costs of industrial research projects and 25 percent for pre-competitive development projects and the firms have to make up the remaining funding, the paper further developed the research by examining whether the firms capital value impacts the effectiveness of the program. To do this, this paper divides the firms into two groups based on the firms pre-program total capital value–the High Capital group, whose pre-program total capital (one year before the program) is above the median, and the Low Capital group, whose pre-program capital is below the median.

The paper then tries to select the ideal bandwidth for further comparison. The paper analyzed the mean differences of firms' pre-program capital, pre-program assets, and pre-program investment for the full sample, wide window (50% of firms in the right and left of the cutoff threshold), and narrow window (35% of firms in the right and left side of the threshold) in both the high capital and low capital group. The paper finds that only the narrow window groups do not display a significant mean difference between the control and treatment groups for both high capital groups and low capital groups over the three matrices. Therefore, the control and treatment groups are considered as similar enough for comparison, and the narrow window was determined for further analysis.

The mean difference results are consistent with the previous speculations and the RDD literature. The selection should ensure that the control and treatment groups have similar qualities. And only samples selected close from the cutoff (score = 75) can be considered as similar. We can speculate that a firm with a score of 30 may be really different from a firm with a score of 90 since the score was given based on their potential to invest in R&D. Therefore, a bandwidth too large would undermine the data's ability to display the real difference between the control and treatment group. On the other hand, the sample size for the two groups needs to be large enough to maintain the statistical power of the analysis. This is why the paper chose multiple bandwidths to analyze and landed on the narrow window bandwidth.

The paper then used the same model from the Bronzini and Iachini paper to examine the effectiveness of the program on firms' R&D outlays to the High Capitial and Low Capital group. The paper finds that for firms with low capital, the coefficients for the treatment variable T_{i} are significant in the total investment/pre-program sales model and the intangible investment/pre-program sales model but not on the tangible investment/pre-program sales model. This suggests that for firms with relatively low capital reserves, the subsidy from the program was effective in supporting their innovation activities and allowed them to have the financial capacities to invest in R&D activities. Since the co-efficient for total investment/pre-program sales model and the intangible investment/pre-program sales are significant, these firms likely tend to use the subsidies for intangible intellectual properties expenditures, such as patents and copyrights.

The paper finds that for firms with high capital reserves, the coefficients for the treatment variable T_{i} are not significant in neither the total investment/pre-program sales model, the intangible investment/pre-program sales model nor the tangible investment/pre-program sales model. This suggests that the program was not the determining factor for the R&D investment increase for firms with high capital reserves. In other words, the program was not an effective tool to support a firm's R&D investment.

## 5.2 Model and data limitation

### 5.2.1 Variable Selection

One major limitation of this research is the representativeness of the outcome assumes. As discussed in 2.2 Variable Selection section, the dependent variables, namely total investment/pre-program sales, tangible investment/pre-program sales, and intangible investment/pre-program sales, are selected to represent the firms' R&D investment. However, these variables are not the perfect representation of firms' R&D investment. The tangible assets were chosen as a representation of the firm's R&D investment because a large share of firms' expenditures for innovation activities (about 40%) is covered by tangible assets. Therefore, Bronzini and Iachini assumes that the net change of the tangible asset could somewhat represent the firm's R&D investment. However, this may introduce noise to the data and undermine the internal validity of the analysis since tangible assets do not cover an average of 60% of Italian firms' expenditure for innovation activities. What's more, the 40% of firms' expenditure for innovation activity is an average statistic for the Italian firms, the data may not be representative of the 357 firms chosen for this paper. Within the 357 firms, there may be nuances and not all firms' innovation expenditure is spent on acquiring tangible assets. This will decrease the internal validity of the research.

The same issue also exists in the use of intangible investment/per-program sales to represent R&D investment. Since the detailed breakdowns of the firms' intangible asset costs are not available for all companies in the dataset, Bronzini and Iachini selected five firms participating in the program with the largest intangible assets (assuming that their data quality would be the best) to study their intangible asset structure. They found that about 66% of the total intangible assets cost are expenditures eligible for the subsidy (e.g.R&D, patents, software, and licenses), therefore, the use of intangible assets as an outcome variable in the analysis is reasonable. However, this variable selection method may reduce the internal validity of the study as (1) the five firms with the largest intangible assets are not randomly selected and may not be representative of the 357 firms. Firms with the largest intangible assets could be firms that are larger in size or more high-tech in their products. (2) Of the five firms selected, 34% of their intangible asset costs are items that are not reimbursable by the program, including goodwill and costs for ongoing intangible assets, which are not related to R&D spendings. The use of intangible assets as the outcome variable means that irrelevant costs increase may be counted as an increase in R&D spending.

Even if all costs for tangible and intangible assets are expenditures that could be reimbursed by the program, the total investment/pre-program sales variable could not capture all R&D investment by the firms. This is because the expenditures for employment of researchers, contracts with research centers, consulting, and feasibility studies may not be captured in either the tangible or intangible assets, therefore, the sum –total investment would not represent these investment items. This paper did not include these variables because there is no clear breakdown in the firms' financials for these items, and unlike expenditures in tangible and intangible assets, the R&D-related expenditure may account for only a small portion of the expenditure. It is not reasonable to deem labour costs and service costs as fair representations of the firm's R&D investments.

Although the limitations for the outcome variables are impeding the internal validity of the analysis, this paper does recognize that the lack of data in studies of firms' innovation investment and expenditure is a prevalent issue. The financial data provided by Bronzini and Iachini is still valuable and worth studying with the caveats mentioned above.

### 5.2.2 Model Selection

This paper used the RDD model to conduct the analysis. Two factors may affect the internal validity of the model. First, the paper used a linear regression model to test the relationship between the variables. However, the data may not fit the first-order polynomial model. The original paper estimates up to a third-order polynomial model on the full sample and found that the AICs for the simpler models are more accurate for the full sample data. Since this paper divided the data into two groups by the pre-program total capital, the same model may not be the best model to explain the relationship.

The other influencing factor is the lack of control variables in the model. Although the paper recognized that the difference in firms' pre-program total capital may influence firms' R&D investment capacity post-program and separated the data into two groups to analyze the program impact in different scenarios, the paper could not control the other factors that might influence firm's investment in R&D, including the size of subsidies firms receives, firm's innovation strategy (Knott and Posen 2009), firm's exposure to foreign markets (Peters, Roberts, and others 2018), and firm's sector (Ehie and Olibe 2010), etc. Due to the lack of relevant data, it is impossible to build these factors as covariates of the RDD model in this paper. This may diminish the internal validity of the model.

### 5.2.3 External Validity

This study is based on the data on 357 industrial firms in the year 2005 with investment projects in Elimlia-Romagna, Italy. This is a relatively specific sample. Since no data on other programs in other places of the world in other periods were represented in the data sample, the results of the study cannot be used to draw inference to the effectiveness of other public subsidy programs, or similar programs in different regions and/or different time.

## 5.3 Future area of study

Based on the limitations in the previous subsections, future work to address these limitations will contribute to the research of the effectiveness of the public innovation subsidy programs. Some area for consideration includes: (1) Find more accurate data variable to represent the change in R&D investment in a firm to increase the internal validity of the RDD model. (2) Consider controlling the possible influences to the effectiveness of the innovation programs by leveraging relevant data and building relevant covariates into the analysis model. (3) Explore the fitness of the model by increasing the level of the polynomial of the model to determine the model that works best for the data. (4) Conduct RDD analysis to data for different government programs in different regions and different periods to increase the external validity of the analysis.

# Reference:


Arel-Bundock, Vincent. 2021. *Modelsummary: Summary Tables and Plots for Statistical Models and Data: Beautiful, Customizable, and Publication-Ready*. https://CRAN.R-project.org/package=modelsummary.

Bronzini, Raffaello, and Eleonora Iachini. 2014. "Are Incentives for r&d Effective? Evidence from a Regression Discontinuity Approach." *American Economic Journal: Economic Policy* 6 (4): 100–134.

Chang, Winston. 2018. *Gcookbook: Data for "r Graphics Cookbook"*. https://CRAN.R-project.org/package=gcookbook.

David, Parthiban, Jonathan P O'Brien, and Toru Yoshikawa. 2008. "The Implications of Debt Heterogeneity for r&d Investment and Firm Performance." *Academy of Management Journal* 51 (1): 165–81.

Ehie, Ike C, and Kingsley Olibe. 2010. "The Effect of r&d Investment on Firm Value: An Examination of US Manufacturing and Service Industries." *International Journal of Production Economics* 128 (1): 127–35.

Gertler, Paul J, Sebastian Martinez, Patrick Premand, Laura B Rawlings, and Christel MJ Vermeersch. 2016. *Impact Evaluation in Practice*. The World Bank.

Knott, Anne Marie, and Hart E Posen. 2009. "Firm r&d Behavior and Evolving Technology in Established Industries." *Organization Science* 20 (2): 352–67.

Muscio, Alessandro. 2014. *Regional Innovation Monitor Plus*. https://ec.europa.eu/growth/tools-databases/regional-innovation-monitor/sites/default/files/report/140120_RIM%20Plus_Regional%20Innovation%20Report_Emilia-Romagna.pdf.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Peters, Bettina, Mark J Roberts, and others. 2018. "Firm r&d Investment and Export Market Exposure." National Bureau of Economic Research.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

StataCorp. 2003. *Stata Software*. https://www.stata.com/.

Statistica, Istituto Nazionale di. 2010.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, and Evan Miller. 2021. *Haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files*. https://CRAN.R-project.org/package=haven.

Zhu, Hao. 2020. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.