Are R&D Subsidies Effective? The Effects of Industry Competition

1. Summary

The main objective of the paper is to analyse the effects of competition intensity on public R&D subsidies effectiveness. The author uses threshold regression that divides the sample of Chinese firms into classes based on their degree of competition face by a firm, and then s/he estimates the differing effects of subsidies on R&D expenditures among these classes. Results suggest nonlinear threshold effects of competition on R&D effectiveness.

2. Detailed comments

The subject of the paper is very interesting and I certainly do recommend further studies in the field. However, the paper itself has certain drawbacks, some of which must and some of which should in my opinion be addressed.

2.1 Major comments

The specifications of the empirical model on P. 5 are not accurate for several reasons:

There is not any explanation provided for the subscripts i and t. Please say what do i and t refer to in the model.

You include control variables in the regression by adding $\sum \beta_i \, Control_{i,t}$ to the model. The issue here is in the subscript i. Initially, the subscript i is used to indicate firm i (e.g. in $R\&D_{i,t}$) and here it is used as a subscript of the coefficient. You can address this issue by changing i into another subscript or you can simply include $\beta_x X_{i,t}$ and say that $X_{i,t}$ is a vector of control variables.

The subsidy variable $Subsidy_{i,t}$ has the subscript i which refers to the firm level. However, in the explanation provided below the model, subsidy appears as a variable that is given at the industry level. "Subsidy denotes instrument variable, which is defined as industry mean of natural logarithm of individual firms' R&D subsidies." Therefore, subsidy should have a subscript j, that refers to the industry level. However, I view that the natural order to present the results is to show first the ones that do not correct for self-selection bias, and then move to the results with subsidy proxy (industry mean) that corrects for the self-section bias. Accordingly, you may leave the subsidy as $Subsidy_{i,t}$ this in the regression as your baseline model.

I do not understand why you do the following: “For robustness, the natural logarithm for the total subsidies allocated to industries is also employed as instrument variable.” For me this is not a robustness check, but rather the same thing as you have done before. That is, if you average over all firms in one industry, then you should get the industry average.

You mention that your econometric model is based on Clausen (2009) who uses instrumental variables (IV) regression. It is not clear to me whether your regression model is the reduced form or the second stage regression. It appears to be a pure reduced form, but it is not clearly written in the text. Please mention what it is. Also, please comment briefly on how your model differs from Clausen’s (2009). But again, please consider what I have written earlier when addressing this point: “However, I view that the natural order to present the results is to show the ones that do not correct for self-selection bias first, and then move to the results with subsidy proxy (industry mean) that corrects for the self-section bias. Accordingly, you may leave the subsidy as $Subsidy_{i,t}$ this in the regression as your baseline model.”

There is not a formal discussion of why the industry mean of natural logarithm of individual firms' R&D subsidies could correct for the self-selection bias. It might only transfer the endogeneity problem to the industry level. For example, government may pick industries where innovation is higher. Therefore, the proposed instrument may not correct the bias.

My understanding is that you are running a threshold regression following Hensen (1990) and that you are trying to estimate the effect of subsidies on R&D expenditures, based on the value of competition. Accordingly, the econometric model should look as follows:

$$R\&D_{i,t} = \beta_1 + \beta_2 Subsidy_{j,t} I\left(Com_{i,t} \leq \alpha_1\right) + \beta_3 Subsidy_{j,t} I\left(\alpha_1 < Com_{i,t} \leq \alpha_2\right) + \beta_4 Subsidy_{j,t} I\left(Com_{i,t} > \alpha_2\right) + \beta_x X_{i,t} + \varepsilon_{i,t}$$

In the regression Tables 4 and 5, on p. 14 and p. 15 respectively, there are three different coefficients for each level of competition, which are:

$$I\left(Com_{i,t} < \alpha_1\right), I\left(\alpha_1 < Com_{i,t} < \alpha_2\right) \text{ and } I\left(Com_{i,t} > \alpha_2\right)$$

These should rather be:

$$I\left(Com_{i,t} \leq \alpha_1\right), I\left(\alpha_1 < Com_{i,t} \leq \alpha_2\right) \text{ and } I\left(Com_{i,t} > \alpha_2\right)$$

1.1 Minor comments

The quality of English is not bad but could be further improved. More appropriate use of academic language would enhance of the paper's quality, which is good overall. I include a few examples below. However, I suggest re-reading the paper carefully.

On p.2, "In this paper, I expand existing literature**s** ….."-→ correction: "In this paper, I expand **the** existing **literature**…."

On p.2, " Up to date, little attention……. On R&D **subsides**…." Subsidies is misspelled.

Inappropriate use of the possessive *s* in many places in the paper. For example, second level's regression, firms' fixed effect…..

The Herfindahl index is misspelled throughout the paper.

On p. 3, "After **these cleaning**, I **employ** a balanced panel **for** 901 publicly traded firms….". I suggest rewriting it as follows: "I ***use*** a balanced panel ***of*** 901 publicly traded firms. " Also, "these" is plural, while "cleaning" is singular. Better would be: "After these cleaning steps…."

The headings are inconsistent. Lower case letters are used most of the time, except for section 3. Empirical Results. It is fine to use either, but just keep the consistency.

On p. 3 in Section 2.1 Dependent variable. *"I use internal R&D expenditures, which subtract government subsidies from total R&D expenditures, to measure the amount of private R&D spending."* Is not that repetition to what has been mentioned earlier? *"The first one is firms' internal expenditures on R&D, which subtract government subsidies from firms' total R&D expenditures (Clausen, 2009)."*

On p. 5, Section 2.3 Control variables. *"Besides, I also control for firms' financial leverage…"* "Besides" could be removed from the sentence.

The names of the tables start with small letters throughout the text. It would be better to start the names of the tables with a capital letter. For instance, on p. 5, line 9 of section 2.3, it is better to replace “table 1” with “Table 1”. Similar issues are found in the rest of the paper. On p. 6, line 7 of section 3.1, it is better to change “table 3” to “Table 3”. On p.8, “table 4” and on p.9, “table 3” and “table 2”.

On p. 9, Section 3.3 Robustness tests. However, the **Herfidahl** index **doesn't** show any cut-point when considering the relationship between subsidies and firms' total R&D spending in column (6) of **table** 2. The **Herfidahl** is misspelled. Please write the full words, not **doesn’t**.

It is advisable to use “and” instead of “&” when you cite the name of the authors throughout the paper.

There are few minor mistakes in the references list to have a look at. In specific, “R&D” is written as “r&d” in many instances in the references list. Also for the consistency in your list of references, you need to change Lee, C.-Y. (2011) to Lee, C.Y. (2011).

Overall, the topic is very interesting. However, the paper requires further improvements and the empirical model needs more clarification.