Data sources and procedures to replicate the results for Thomson, R, The Effect of R&D tax incentives. Review of economics and statistics

Analysis undertaken using STATA version 13.1

Files required to generate the result include 2 STATA .do code and 6 data files (6 files). These are described and outlined below.

(A) .do files (code)

01_Prepare_data.do. This file takes as input all raw data and produces the final regression sample.

02_regression_results. This data produces the regression results reported in table 5 of the paper.

nameclean.ado - an .ado file (stata function) that is called to clean string data including country and industry names, convert to lowercase etc. This needs to be placed in the appropriate directory and executed in Stata before the main two codes will run properly.

Each of the .do files includes comments and documents to explain each step.

Outputs:

Running the do files in order 01 then 02 produces three tables in the form of excel spreadsheets:

Table 5: the main regression results reported in the paper.

Results unreported.xls shows results pertaining to other robustness checks discussed in the paper, but not reported including dynamic structure, additional control variables etc.

Results unreported 2.xls and Results unreported 3.xls report results for the main variables systematically dropping one country at a time and then one industry at a time, respectively.

(B) Raw (original) data

1. R&D tax data.xlsx.

This file contains the national level R&D tax policy variables including the calculations used to derive these. Source: authors calculations. The raw data are described in detail in Thomson (2013), as cited in the manuscript. These include the tax price of each expenditure component which are weighted by expenditure shares in the analysis.

2. berd_cost_industry.csv

R&D expenditure shares by type (capital / labor etc) used to produce the industry level tax variable. Source: OECD 2011a, R&D Data by Type of Expenditure by Industry, Business Enterprise R-D Expenditure by Industry and by Type of Cost, accessed 15/11/2011.

3. berd_by_ind_by_funds.csv

Dependent variable for regression is R& D financed by industry. Source OECD Business Enterprise R&D Expenditure by Industry and by Source of Funds,; Accessed on 17 Feb 2012

4. stan value added.csv.

Data on value added by sector (control variable) Source: OECD 2011b, Structural Analysis Statistics, Value Added by Industry, accessed 15 Nov 2011.

5. msti.csv

Data source for additional control variables used for robustness checks. Not used for results reported in any table in the manuscript. Source: OECD Main Science and Technology Indicators. Accessed on 24 Feb 2012

6. oecd_gdp.csv OECD main national accounts data from stat.oecd accessed 16 Apr 2015

The variables used in the regression analysis are :

Lrdfind – Natural logarithm of business expenditure on R&D financed by Industry.

Livadollars – Natural logarithm of industry value added in constant PPP, constant prices with base year 2000.

lf_b – Natural logarithm of the tax price of R&D (industry specific as described in the paper).

The panel variables are denoted (i,t) where *i* is an industry-country dummy and *t* is year.