

The effect of R&D expenditure on the number of patents

1. CUSIP - Compustat's identifying number for the firm (Committee on Uniform Security Identification Procedure number).
2. SCIENTIFIC - Dummy equal to one for firms in the scientific sector.
3. LOG-CAPITAL - The logarithm of inflation adjusted book value of capital in 1972
4. SUMPAT - The sum of patents applied for between 1972-1979
5. LOG-RD - The logarithm of R and D spending during the year (in 1972 dollars)
6. PATENTS - The number of patents applied for during the year that were eventually granted.
7. Year - Year

The objective of our paper is to look at elasticity of patenting with respect to R & D in order to examine the effect of R & D expenditure on the number of patents established by firms. The nature of dataset is a panel data for years 1972 to 1979.

For the OLS model the dependent variable in our analysis is the patents corresponding to number of patents for each firm in a given year and our independent variables are Log R & D spending during year, LOG-CAPITAL and SCIENTIFIC. We are dealing with balanced panel data in which there are equal number of observations for each firm during the time frame.

We are going to examine this effect both as a cross-sectional study between firms in a given year and also as a panel study to examine this effect during the time range we have. As for some firms the number of patents is equal to zero in some years, we are going to use sumpat variable to get a sum value of patents for each firm.

Below we provided the general summary statistics and description of dataset along with summary statistics specifically for panel dataset. Y using *xtset* and *xtsum* commands. We also provided the summary statistics for each time frame individually by using *bysort* command.

There are 8 time-spans from 1972 to 1979. There are 642 observations of firms in each time span and 5136 observations in total.

Contains data
Observations: **5,136**
Variables: **7**

Variable name	Storage type	Display format	Value label	Variable label
cusip	float	%9.0g		
scientific	float	%9.0g		
logcapital	float	%9.0g		
sumpat	float	%9.0g		
logrd	float	%9.0g		
patents	float	%9.0g		
year	float	%9.0g		

Sorted by: **cusip year**
Note: Dataset has changed since last saved.

. summarize

Variable	Obs	Mean	Std. dev.	Min	Max
cusip	5,136	49461	28680.02	170	99850
scientific	5,136	.3691589	.482624	0	1
logcapital	5,136	3.547407	2.062251	-2.0804	9.66626
sumpat	5,136	210.4642	533.3099	0	6734
logrd	5,136	.8049331	1.953528	-4.48488	7.06524
patents	5,136	26.30802	67.79056	0	906
year	5,136	1975.5	2.291511	1972	1979

. xtset cusip year

Panel variable: **cusip** (strongly balanced)
Time variable: **year, 1972 to 1979**
Delta: **1 unit**

. xtdescribe

cusip: **170, 181, ..., 99850** n = **642**
year: **1972, 1973, ..., 1979** T = **8**
Delta(year) = **1 unit**
Span(year) = **8 periods**
(cusip*year uniquely identifies each observation)

Distribution of T_i: min 5% 25% 50% 75% 95% max
8 8 8 8 8 8 8

Freq.	Percent	Cum.	Pattern
642	100.00	100.00	11111111
642	100.00		XXXXXXXX

.

Variable		Mean	Std. dev.	Min	Max	Observations	
cusip	overall	49461	28680.02	170	99850	N =	5136
	between		28699.59	170	99850	n =	642
	within		0	49461	49461	T =	8
scient~c	overall	.3691589	.482624	0	1	N =	5136
	between		.4829533	0	1	n =	642
	within		0	.3691589	.3691589	T =	8
logcap~l	overall	3.547407	2.062251	-2.0804	9.66626	N =	5136
	between		2.063658	-2.0804	9.66626	n =	642
	within		0	3.547407	3.547407	T =	8
sumpat	overall	210.4642	533.3099	0	6734	N =	5136
	between		533.6737	0	6734	n =	642
	within		0	210.4642	210.4642	T =	8
logrd	overall	.8049331	1.953528	-4.48488	7.06524	N =	5136
	between		1.926499	-3.856315	6.948887	n =	642
	within		.3315607	-2.496041	3.393241	T =	8
patents	overall	26.30802	67.79056	0	906	N =	5136
	between		66.70922	0	841.75	n =	642
	within		12.3088	-207.442	198.058	T =	8
year	overall	1975.5	2.291511	1972	1979	N =	5136
	between		0	1975.5	1975.5	n =	642
	within		2.291511	1972	1979	T =	8

In this summary statistics we can observe that the std for some variables such as cusip, scientific, logcapital, sumpat, and year within the duration of study is equal to zero. This makes sense as these variables are constant for each firm within the proposed time frame. For example sumpat is a number corresponding to total sum of patents for each firm during the time range which is constant for each firm. However, we observe that sumpat between firms has std of 533.3099 with mean value of 210.4642. Also, the number of patents between firms has std of 67.790 with mean of 26.30. Furthermore, the std for logrd between firms is 1.9535 with mean of 0.804.