

The Economics of Innovation and Intellectual Property

Chapter 6: Exercise 2

Bronwyn H. Hall & Christian Helmers
2024

In this exercise, you will reproduce some estimates from Tables 5 and 6 in Hall, Jaffe, and Trajtenberg (2005), Market Value and Patent Citations, *Rand Journal of Economics*, 36: 16-38. Those estimates were obtained using nonlinear least squares estimation, but for simplicity, we use the linear approximation to the equation in this exercise. [**Note:** you can carry out the statistical analysis using your preferred statistical software (e.g. Stata, R, MS Excel, Python, etc.)]

The data for this exercise are available in the files chapter6_2_data.dta (for Stata) and chapter6_2_data.xlsx (MS Excel).

The data are for an unbalanced panel of 1,982 publicly traded U.S. manufacturing firms that patent at least once during the 1979-1988 period. The questions below will make use of the following 12 variables (there are many more on the file which you are welcome to ignore):

Variable	Description
citepat	Citation stock/patent stock ratio
drnd	Dummy for zero R&D
id	Numeric id for firm
ind	Quasi-two-digit industry of the firm
ka	K/A , R&D stock/capital stock ratio
logq	Logarithm of Tobin's q
netcap	A , Net capital stock (end of year, \$USM)
patrnd	Ratio of patent stock to R&D stock in \$USM
selfpat	Self citation stock/patent stock ratio
tech	Broad technology class indicator for firm's industry (1-6)
year	4-digit Year

The broad technology-industry classes are the following:

Tech	Industry
1	Drugs & medical instruments
2	Chemicals
3	Computers & instruments
4	Electrical
5	Metals & machinery
6	Miscellaneous other industries

In Chapter 6, the market value model including R&D capital is given by the following:

$$\log Q_{it} = q_t + (s - 1) \log A_{it} + sg \frac{K_{it}}{A_{it}} + e_{it} \quad (1)$$

where $i = 1, \dots, N$ and $t = 1, \dots, T$, i is the firm and t indicates time. Q is the ratio of market value to book value, A is tangible capital stock, and K is R&D capital, all in logs. Each year has a separate intercept, denoted q_t , which will be included in the equation as time or year dummies.

1. Report sample statistics (number of observations, mean, median, standard deviation, min, max) for the key variables $\log q$, $\log A$, K/A , and $drnd$.
 - (a) What does the median of $\log q$ tell us about the valuation of capital?
 - (b) What is the range of K/A values? Does the approximation of $\log(1 + gK/A) \cong gK/A$ seem valid to you for these data?
 - (c) What fraction of the observations have zero R&D stock? We will control for this in what follows.
2. Compute the means and medians of K/A by tech (i.e. industry), and by year. Do they differ across industries in the ways you expect? How does the ratio evolve over time?
3. Given what we learned in 1(b) above, now restrict the sample for estimation to observations with K/A less than 0.4.
 - (a) Estimate the above equation by OLS (ordinary least squares). Don't forget to control for the year, as we expect average q to differ across time, based on our statistics in (2). Also, remember to include a dummy for those firms with zero R&D, to ensure that they are allowed to have their own mean $\log q$. Report the results.
 - (b) Derive the estimated g from the above equation and interpret what the value implies.
 - (c) Now add the tech sector dummies to the equation. How do the results change?

4. Now add the ratio of patent stock to R&D stock to the equation in 3(a) and re-estimate. What does the result tell you? How would you interpret this coefficient?
5. Now add the ratio of the citation to patent stock to the equation in (4) and re-estimate. What does the result tell you? How would you interpret this coefficient?