Who Invents, When and Where?

Preliminary draft and results, please do not quote

15.3.2009

Otto Toivanen and Lotta Väänänen*

ABSTRACT

We study individual- and firm level determinants of invention using a dataset on U.S. patents' Finnish inventors and their employers. We find that females, entrepreneurs, and unemployed individuals are less inventive. We find significant and large differences between different fields and levels of education: engineering education has a positive significant coefficient at all levels of education, with the magnitude increasing with the level. At the doctorate level, also the coefficients for the fields of natural sciences, and health and welfare are large and significant. Our estimates of the ageing effects vary between the different specifications used, highlighting the difficulty in obtaining consistent estimates of the ageing effects. With regard to the question of where one invents, our results show that firm size is positively associated with the propensity to patent, even at the level of an individual. We also find that R&D spending per R&D employee is positively associated with the propensity to patent.

KEYWORDS: ageing effects, education, innovation, inventors, patents, productivity, scale effects

JEL codes:

---

* Toivanen: HECER, PO Box 17, FIN 00014 University of Helsinki, Finland. Email: otto.toivanen@helsinki.fi. Väänänen: Helsinki School of Economics and HECER, Department of Economics, P.O. BOX 1210, 00101 Helsinki, Finland. E-mail: lotta.vaananen@hse.fi. We would like to thank the Yrjö Jahnsson Foundation for financial support. Väänänen also thanks the FDPE for financial support. The usual caveat applies.

1 Introduction

Invention, defined as activity directed toward the discovery of new and useful knowledge about products and processes, is one of the most important phases of the growth of civilization. Yet it is one of the least understood. Who engages in an inventive activity, why, when, and how? With these words, Jacob Schmookler defined the research agenda to which this paper belongs - half a century ago (Schmookler 1957, pp. 321). He went on to say that the question was starting to receive attention even if the amount of work was "lamentably small". While there has been progress since his remarks, in particular the recent use of the inventor information in the NBER patent data (see Trajtenberg et al. 2006), these questions remain largely unanswered. Our objective in this paper is to contribute to the fulfillment of the agenda by answering the questions posed in the title of the paper: "Who invents, when, and where?"

We study one of the central questions in the economics of innovation, the propensity to patent (see e.g. Scherer 1967, Hall, Hausman and Griliches 1984), but at the unusual level of the individual inventor (see also the work of Kim, Lee and Marschke 2007). We use a dataset where USPTO patents and their inventors from the NBER patents and citations data file (Hall, Jaffe Trajtenberg, 2001) are linked to Finnish employer-employee data containing detailed information on the individuals and the employers from 1988 to 1999.[1] We study how individual and employer characteristics affect the propensity to patent. We attempt to identify ageing effects to look at the question of life-cycle productivity (see Levin and Stephan 1991), acknowledging that it is

[1] Finland is one of the few countries that has successfully transformed its inventive capacity in the last few decades. In terms of patenting, the change is on par with that experienced by Israel, Taiwan and South Korea (Trajtenberg 2001). Toivanen and Väänänen (2008) use the same data to study returns to inventors.

difficult to disentangle age-, cohort-, and year effects (see Hall, Mairesse and Turner 2005). We take two approaches to try to deal this problem: first, we group (ad-hoc) cohorts; second, we replace calendar time effects with measures of technology- and R&D-intensity in the economy each year. We also look at the fields and levels of education of the inventors, and investigate whether age-profiles vary depending on education.

Our descriptive statistics show, that relative to the population of working-aged Finns, inventors are less likely to be female (5-10%). The average age of inventors is 41 years. Not surprisingly, inventors are more likely to be employed (and less likely to be students, retired, or unemployed). Inventors also differ from the population with respect to their education: inventors are much more likely to have a high-school diploma, a masters degree or a doctorate, and they are more likely to come from the fields of engineering or natural sciences. Finally, in terms of their occupations, inventors are much more likely to come from the occupational groups of professionals or managers.

The results from our pooled OLS estimations of inventive productivity lend further support to what our descriptive comparisons show: that females, entrepreneurs, and unemployed individuals are less likely to invent. We also find significant and large differences between different fields and levels of education: engineering education has a positive significant coefficient at all levels of education, with the magnitude increasing with the level. At the doctorate level, also the coefficients for the fields of natural sciences and health and welfare are large and significant, while also resources and services are positive and significant. These results are partial correlations that do not control for the potential endogeneneity of education; identifying the causal effect of education using instrumental variables is the topic of another on-going paper.[2]

Our estimates of the coefficient of age (and age squared) vary between the different specifications used, as does the estimated age of peak productivity, highlighting the difficulty in obtaining consistent estimates of the ageing effects.

With regard to the question of where one invents, our results echo the common finding from previous studies that firm size is positively associated with the propensity to patent, even at the level of an individual. We also find that R&D spending per R&D employee is positively associated with the propensity to patent.

The amount of work on private sector inventors is limited by the availability of data, thus our descriptive analysis provides new insights into how individual characteristics, such as gender, age, and education are associated with the propensity to patent. In previous work, Giuri et al. (2007) report characteristics of European inventors based on the PatVal survey: only 2.8% are women, mean age is 45, and 77% have a university degree (26% a doctorate). Khan and Sokoloff (2004) demonstrate that an elite background was not a determinant of being able to become a great inventor in the 19th and early 20th centuries. Kim, Lee and Marschke (2007) study the productivity of inventors in U.S. pharmaceutical and semiconductor firms. They find that productivity increases with firm size (measured by R&D, sales or number of employees) even after controlling for inventor specific unobserved heterogeneity. The capital-labor ratio affects individual inventiveness positively. They also find that experience affects inventiveness positively but nonlinearly, with the effect turning negative at high levels of experience.

[2] In ongoing work, we estimate the causal effect of education on invention using variation over time and geographically in the possibility to obtain an (either college or university) engineering degree (in the 1970s and 1980s).

We are able to look at the question of how employer and employee characteristics interact to affect inventive productivity. It is relevant, given not only the original Schmookler worry – that the then experienced demise of the independent inventor could lead to a slow-down in the inventive process – but also from today's perspective as some industries are experiencing a reversal of this phenomenon (e.g. biotech), with invention primarily taking place in small start-ups.

There is a relatively large literature on research productivity of academics. The "who invents" and "when one invents" questions have been addressed in the inventive work of Levin and Stephan (1991) and Jones (2007). Levin and Stephan (1991) study U.S. academic scientists and find life-cycle aging effects. Jones (2007) studies the "great inventors" and Nobelists and finds that the age at which they make their main contributions has been increasing. Carayol and Matt (2006) study academics of the French Louis Pasteur University and find that the size of the lab has a negative effect on output. Kelchtermans and Veugelers (2006) study Catholic University of Leuven scientists. They find that rank, gender, hierarchical position and past performance are all important explanatory factors in determining scientific output. Wuchty, Jones and Uzzi (2007) show the rising importance of teams in academic research and in patenting.

The rest of the paper proceeds as follows. Section 2 describes the data and presents a comparison between inventors and non-inventors. Special attention is paid to the differences in levels and type of education. Section 3 presents the empirical framework. In section 4 we present the results and in section 5 the conclusions.

2 Data and descriptive analysis

2.1 Data

Our data comes from three sources. Information on inventors and USPTO patents comes from the NBER patents data base described in Hall, Jaffe Trajtenberg (2001). This data is matched to the Finnish Linked Employer-Employee data of Statistics Finland (FLEED). FLEED is described in Korkeamäki and Kyyrä (2000) and the matching process in Toivanen and Väänänen (2008). We also link data from the Finnish R&D survey to the companies in FLEED. The R&D survey contains the majority of the firms that do R&D and patent, and it includes variables describing the firm's R&D expenditures and personnel. This we use to examine the question of "where" one invents.

To identify the Finnish inventors from the NBER patent data, the information contained in the patent records (name of individual, municipality in which the individual resided at the time) was used at the Statistics Finland to search the Finnish Population Information System (FPIS) for their personal identification numbers. These personal identification numbers link the individuals to the records in FLEED, including to their employers as well. When the information in the patent records produced a match with more than one person in the FPIS, we picked the individual whose employer in the FLEED matched the patent assignee in the USPTO data. When this process failed to identify a single individual, we excluded such individuals from our data. Out of 8065 inventor-patent records we were able to match 5905, consisting of 3253 individuals.

For the questions of "who invents and when", our analysis is based on the population of working-aged individuals in Finland. We use a 100 000 random sample from the FLEED, together with our sample of inventors, and we weight the sample to represent the full population. To look at the question of “where” one invents, we restrict our sample to individuals who are employed by firms that are found in the R&D survey.

2.2 Descriptive statistics

Table 1a shows the means for the key variables for inventors, i.e. those individuals who were inventors in a patent application in the given year. Table 1b shows the means for a random sample of the Finnish working-aged population. Only 5-10% of the inventors are female, although this share seems to have been going up slightly over the years. Individuals who have completed their high-school diploma are overrepresented among inventors relative to the population. Similarly, individuals who are employed are overrepresented, while unemployed, students and retired individuals are underrepresented.

Tables 1a and 1b here

Table 2 shows the educational levels for the inventors and for the Finnish working-aged population. Not surprisingly, individuals with a masters degree or a doctorate are much overrepresented among inventors.

Table 2 here

Table 3 shows the educational fields for the inventors and for the Finnish working-aged population. In terms of fields of education, natural sciences and engineering are the fields that are overrepresented among inventors.

Table 3 here

Table 4 shows the occupations for the inventors and for the Finnish working-aged population. Managers, and in particular “professionals” are overrepresented among inventors.

Table 4 here

3 The empirical framework

To maintain some comparability to existing research on the propensity to patent, we employ the “innovation production function approach” that has been the workhorse of firm level analyses at least since the Hall, Hausman and Griliches (1984). We thus estimate equations of the following form:

(1) $Y_i = f(X_i, \theta)$.

Y_i is our output measure (number of patents granted to individual *i*, or citations received by the patents of individual *i*), X_i are possibly time-varying explanatory variables, and θ a vector of parameters to be estimated.

We first estimate the number of patents applied by individual i in year t conditional on variables describing the individual X, cohort C, and time T. These estimations provide descriptive results on the questions of “who invents and when” for the population of working-aged individuals in Finland. We use a 100 000 random sample for each of the years from the FLEED, together with our sample of inventors, and we weight the sample to represent the full population.

When we explore the question of “where” one invents, we restrict our sample to individuals who are linked to information on their employers, i.e. individuals employed at the end of the year, and to firms that are found in the R&D survey. We estimate the effect of firm-level variables on the productivity of an individual using pooled estimations, as well as fixed effects estimations where we can control any individual ability factor that is constant over time (loosing individuals who do not patent over the time period).

4 Results

4.1 Who Invents and When?

To explore the question of "who invents?" we include in our estimating equation variables for gender, nationality (Finnish, foreign), language (Finnish, Swedish, other) dummies for levels and fields of education, and labor market status (employee, entrepreneur, student, unemployed, retired, other). We find that females, entrepreneurs, and unemployed individuals are less likely to invent. We also find significant and large differences between different fields and levels of education: engineering education has a positive significant coefficient at all levels of education, with the magnitude increasing with the level of education. At the doctorate level, also the coefficients for the fields of natural sciences and health and welfare are large and significant, while also resources and services are positive and significant. Figure 1 shows the coefficients on the education dummies from an OLS regression.[3]

Figure 1 here

We address the question of inventive productivity over the life-cycle and try to identify ageing effects. The problem in identifying age-, cohort, and calendar time effects is well known (see Hall, Mairesse, Turner, 2005). We try to deal this problem by replacing calendar time effects with measures of technology- and R&D-intensity in the

[3] The Poisson and Negative Binomial estimators did not converge, which is why we only report OLS results for now.

economy each year. We include total R&D expenditures as well as the value of high-tech exports in each year.

Figure 2 shows the estimated ageing effects from four different specifications with age and its square. The first specification includes calendar year effects together with a quadratic in age (no cohort effects). The second specification includes in addition 4 ad-hoc cohort effects (born before 1950, born in the 50s, born in the 60s, born after the 60s). The third specification includes a full set of cohort-dummies (one for each birth year), but replaces calendar time effects with a measure of total R&D expenditures in the economy. The fourth specification adds another control for calendar time effects, the value of high-tech exports. The estimated ageing effects vary a lot between the specifications, as does the estimated age of peak productivity, highlighting the difficulty in obtaining consistent estimates of the ageing effects. (work in progress).

Figure 2 here

We also investigate how the productivity of people with different levels and fields of education varies with age (work in progress).

4.2 Where?

To analyze the question of where one invents, we restrict our sample to individuals who are employed by firms that are found in the R&D survey. We include variables for the firm size as measured by the number of employees, firm's R&D spending per R&D employee, and the share of R&D workers of total employees. We also include variables for the share of R&D workers with a university education and the share of female R&D workers.

We estimate these using a linear fixed- effects specification.[4] In line with results from previous studies, firm size is positively associated with the propensity to patent, also at the individual level. We also find that firm's R&D spending per R&D worker is positively associated with the propensity to patent. Tenure with current employer has a non-linear relation with productivity, first increasing it, and eventually decreasing it.

Table 5 here

5 Conclusions

We study who invents (patents) and where using a matched data set on Finnish inventors of U.S. patents and their employers covering the period 1988-1996. These questions are central to further our understanding of the inventive process, which is at the heart of advances in economic growth.

We find significant and large differences between different fields and levels of education: engineering education has a positive significant coefficient at all levels of education, with the magnitude increasing with the level. At the doctorate level, also the coefficients for the fields of natural sciences, and health and welfare are large and significant. These results are partial correlations that do not control for the potential endogeneneity of education; identifying the causal effect of education using instrumental variables is the topic of another on-going paper. Our estimates of the ageing effects vary between the different specifications used, as does the estimated age of peak productivity, highlighting the difficulty in obtaining consistent estimates of the ageing effects. We also find that females, entrepreneurs, and unemployed individuals are less inventive. With

[4] We had trouble getting the fixed effects Poisson estimations to converge.

regard to the question of where one invents, our results are in line with previous studies showing that firm size is positively associated with the propensity to patent, even at the level of an individual. We also find that R&D spending per R&D employee is positively associated with the propensity to patent.

References

Acemoglu, Daron, 2002, Technical Change, Inequality, and the Labor Market, *Journal of Economic Literature*, Vol. 40, Iss. 1, pp. 7-72.

Carayol, Nicolas and Mireille Matt, 2006, Individual and Collective Determinants of Academic Scientists' Productivity, *Information Economics and Policy*, 18, pp. 55–72.

Giuri et al., 2007, Inventors and invention processes in Europe: Results from the PatVal-EU survey, *Research Policy*, 36, pp. 1107–1127.

Hausman, Jerry, Bronwyn Hall and Zvi Griliches, 1984, Econometric Models for Count Data with an Application to the Patents-R&D Relationship, *Econometrica*, Vol 52. No. 4, pp. 909-938.

Hall, Bronwyn, Adam B. Jaffe and Manuel Trajtenberg, 2001, The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools, NBER Working Paper 8498.

Hall, Bronwyn, Jacques Mairesse and Laure Turner, 2005, Identifying Age, Cohort and Period Effects in Scientific Research Productivity: Discussion and Illustration Using Simulated and Actual Data on French Physicists, NBER Working Paper 11739.

Jones, Benjamin F., 2007, Age and Great Invention, mimeo.

Kelchtermans, Stijn and Reinhilde Veugelers, 2006, Top Research Productivity and Its Persistence, CEPR Discussion Paper No. 5415.

Khan, Zorina B. and Kenneth L. Sokoloff, 2004, Institutions and Democratic Invention in 19th-Century America: Evidence from "Great Inventors", 1790-1930, *The American Economic Review Papers and Proceedings*, Vol 94, No.2., pp. 395-401.

Kim, Jinyoung., Sangjoon John Lee and Gerald Marschke, 2007, Research Scientist Productivity and Firm Size: Evidence from Panel Data on Inventors, The Institute for Economic Research, Korea University, Discussion Paper No. 07-08

Korkeamäki, Ossi and Tomi Kyyrä, 2000, Integrated Panel of Finnish Companies and Workers, VATT Discussion Paper No. 226.

Levin, Sharon and Paula Stephan, 1991, Research Productivity Over the Life Cycle: Evidence for Academic Scientists, *The American Economic Review*, Vol. 81, No. 1, pp.114-132.

Pekkarinen, Pekkala and Uusitalo, 2006, Educational Policy and Intergenerational Income Mobility: Evidence from the Finnish Comprehensive School Reform, IFAU Working Paper 2006:13.

Scherer 1967

Schmookler,. J., 1957, Inventors Past and Present, *The Review of Economics and Statistics*, Vol. 39, No. 3, pp. 321-333.

Toivanen, Otto and Lotta Väänänen, 2008, Returns to Inventors, HECER Discussion Paper No. 237.

Trajtenberg, Manuel, 1990, A Penny for Your Quotes: Patent Citations and the Value of Innovations, *Rand Journal of Economics*, Vol. 21, pp.172-187.

Trajtenberg, Manuel, 2001, Innovation in Israel 1968–1997: a Comparative Analysis Using Patent Data, *Research Policy*, Vol. 30, Iss. 3, pp. 363-389.

Wuchty, Stefan, Benjamin F. Jones and Brian Uzzi, 2007, The Increasing Dominance of Teams in Production of Knowledge, Science, Vol. 316, May 2007, pp.1036-1039.

Table 1a. Descriptive stats for the inventors

variable	1988	1989	1990	1991	1992	1993	1994	1995	1996
age	41	41	41	41	41	41	41	41	45
female	0.06	0.05	0.06	0.05	0.07	0.06	0.10	0.07	0.09
high-school diploma	0.65	0.65	0.69	0.65	0.71	0.70	0.73	0.72	0.57
entrepreneur	0.06	0.08	0.06	0.05	0.05	0.05	0.04	0.04	0.06
employed	0.93	0.96	0.97	0.94	0.92	0.94	0.95	0.95	0.82
unemployed	0.00	0.00	0.01	0.02	0.02	0.01	0.00	0.01	0.04
student	0.03	0.00	0.01	0.00	0.01	0.02	0.01	0.01	0.01
retired	0.02	0.02	0.01	0.02	0.04	0.02	0.01	0.02	0.11
finn	0.99	0.99	0.99	0.99	0.98	0.98	0.99	0.98	0.99
finnish	0.88	0.91	0.92	0.91	0.90	0.91	0.92	0.91	0.93
swedish	0.11	0.08	0.06	0.08	0.08	0.07	0.06	0.07	0.05
born<1950	0.58	0.50	0.47	0.46	0.39	0.39	0.36	0.30	0.42
1949<born<1960	0.33	0.40	0.40	0.37	0.39	0.39	0.39	0.38	0.31
1959<born<1970	0.06	0.08	0.11	0.16	0.21	0.21	0.22	0.28	0.25
1969<born	0.01	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01

Table 1b. Descriptive stats for the working-aged population

variable	1988	1989	1990	1991	1992	1993	1994	1995	1996
age	39	40	40	41	41	41	41	41	41
female	0.50	0.50	0.50	0.51	0.50	0.50	0.50	0.50	0.50
high-school diploma	0.20	0.20	0.21	0.21	0.22	0.22	0.23	0.24	0.24
entrepreneur	0.10	0.10	0.09	0.09	0.08	0.08	0.08	0.07	0.07
employed	0.69	0.68	0.66	0.61	0.56	0.52	0.53	0.53	0.54
unemployed	0.04	0.03	0.04	0.08	0.12	0.15	0.14	0.13	0.13
student	0.07	0.07	0.08	0.08	0.09	0.10	0.10	0.11	0.10
retired	0.13	0.14	0.15	0.16	0.17	0.17	0.17	0.17	0.17
finn	1.00	1.00	0.99	0.99	0.99	0.99	0.99	0.99	0.98
finnish	0.94	0.94	0.94	0.94	0.93	0.93	0.93	0.93	0.93
swedish	0.06	0.06	0.06	0.05	0.06	0.06	0.06	0.06	0.05
born<1950	0.50	0.49	0.48	0.47	0.45	0.44	0.42	0.40	0.38
1949<born<1960	0.24	0.23	0.23	0.22	0.22	0.22	0.22	0.22	0.22
1959<born<1970	0.21	0.21	0.21	0.21	0.20	0.20	0.20	0.20	0.20
1969<born	0.05	0.07	0.09	0.10	0.12	0.14	0.15	0.17	0.19

Table 2. Levels of education

	1988	1989	1990	1991	1992	1993	1994	1995	1996
Working-aged population									
upper secondary	36.2	36.1	36.4	36.2	36.0	36.6	37.1	37.8	38.0
lowest tertiary	10.2	10.5	11.2	11.3	11.9	12.1	12.6	12.9	13.3
lower-degree (bachelor)	4.2	4.1	4.2	4.3	4.2	4.4	4.4	4.5	4.6
higher-degree (master)	3.8	3.9	4.0	4.2	4.4	4.6	4.8	5.0	5.1
doctorate	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.5	0.5
unknown	45.4	45.1	43.9	43.7	43.1	42.0	40.8	39.4	38.5
Inventors									
upper secondary	10.7	9.7	8.1	6.6	8.4	8.9	8.3	6.7	12.4
lowest tertiary	11.9	11.0	10.0	10.1	10.8	10.8	8.7	9.1	8.5
lower-degree (bachelor)	14.3	20.3	19.4	21.0	14.9	16.6	16.9	17.0	15.8
higher-degree (master)	35.1	33.6	38.5	43.0	41.8	39.0	42.6	42.7	33.9
doctorate	19.4	20.3	19.4	13.5	20.0	20.0	19.5	19.6	11.8
unknown	8.6	5.2	4.6	5.8	4.2	4.6	4.0	4.9	17.7

Table 3. Fields of education

	1988	1989	1990	1991	1992	1993	1994	1995	1996
Working-aged population									
general	7.0	6.8	6.8	6.7	6.7	6.8	7.0	7.3	7.4
teacher education	1.5	1.5	1.6	1.6	1.6	1.7	1.8	1.8	1.8
humanities & arts	1.6	1.7	1.6	1.6	1.8	1.8	1.8	1.9	1.9
social science & business	9.6	9.8	10.2	10.3	10.9	10.9	11.1	11.3	11.6
natural sciences	0.9	0.9	0.8	0.8	0.9	0.8	0.9	0.9	0.9
technology	18.5	18.4	18.9	18.8	18.9	19.1	19.3	19.7	19.9
agriculture and forestry	3.2	3.2	3.3	3.1	3.1	3.2	3.3	3.2	3.2
health and welfare	5.2	5.3	5.7	5.9	6.0	6.4	6.6	6.9	7.1
services	7.2	7.4	7.2	7.5	7.1	7.3	7.5	7.7	7.7
other or unknown	45.4	45.1	44.0	43.7	43.1	42.0	40.8	39.4	38.5
Inventors									
general	3.3	2.6	2.4	1.6	1.5	2.9	2.8	2.4	2.9
teacher education	0.0	0.0	0.3	0.3	0.0	0.0	0.0	0.0	0.9
humanities & arts	1.2	0.3	1.1	0.8	0.0	0.5	0.4	0.2	0.4
social science & business	2.7	2.6	0.8	1.6	1.2	2.4	0.8	1.6	4.6
natural sciences	11.6	12.6	12.1	10.1	16.7	12.5	16.1	11.9	6.8
technology	66.1	71.0	73.1	76.4	72.2	71.6	69.5	71.5	58.3
agriculture and forestry	2.1	2.6	1.9	1.9	0.3	0.5	1.3	0.6	2.2
health and welfare	3.9	3.2	3.2	1.6	3.9	5.1	4.6	6.9	4.3
services	0.6	0.0	0.5	0.0	0.0	0.0	0.6	0.0	2.0
other or unknown	8.6	5.2	4.6	5.8	4.2	4.6	4.0	4.9	17.7

Table 4. Occupations

Occupation	Population	Inventors
0 Armed forces	0.5	0.0
1 Legislators, senior officials and managers	3.7	11.4
2 Professionals	14.0	67.6
3 Technicians and associate professionals	17.5	15.7
4 Clerks	10.0	0.5
5 Service and care workers, and sales workers	16.0	0.5
6 Skilled agricultural and fishery workers	6.8	0.3
7 Craft and related trades workers	12.1	1.2
8 Plant and machine operators and assemblers	10.3	0.8
9 Elementary occupations	9.1	2.0

Figure 1. Coefficients on education dummies (from OLS regression)

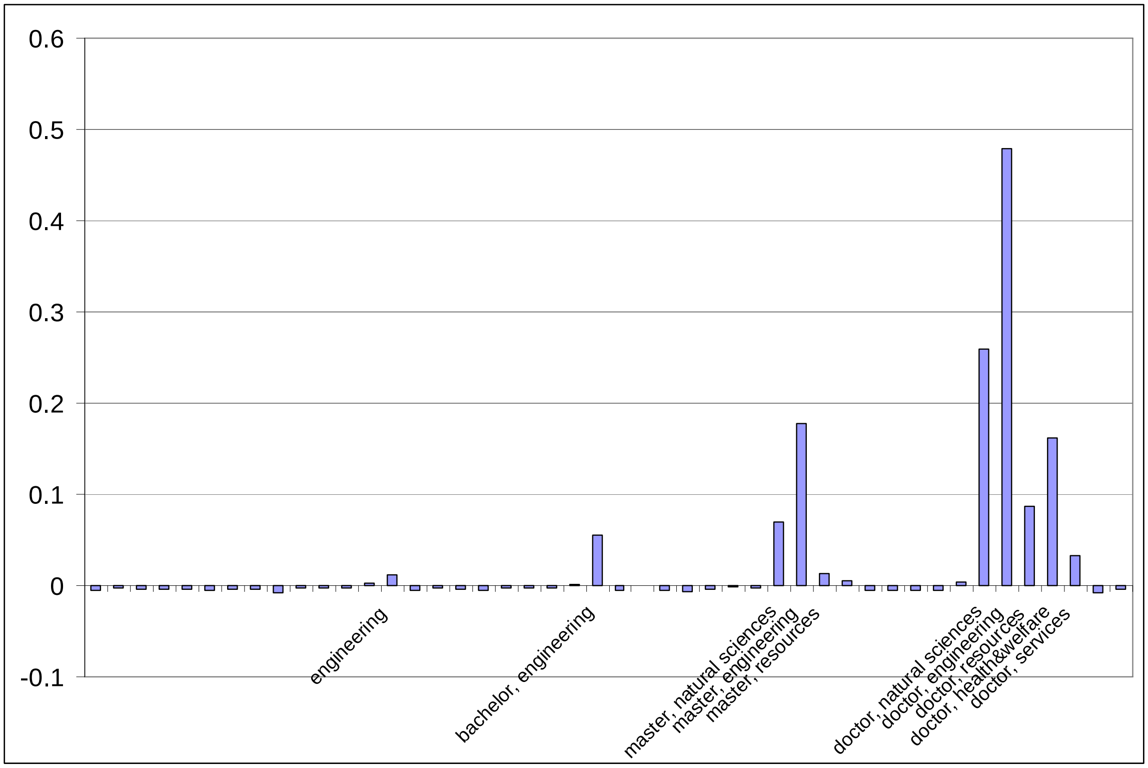

Figure 2. Estimates of ageing effects from different specifications (OLS)