# On the Returns to Invention Within Firms: Evidence from Finland*

Philippe Aghion Ufuk Akcigit Ari Hyytinen Otto Toivanen

January 14, 2018

*– Forthcoming in the AER Papers and Proceedings May 2018 –*

**Abstract**

In this paper we merge individual income data, firm-level data, patenting data and IQ data in Finland over the period 1988-2012 to analyze the returns to invention for inventors and their co-workers or stakeholders within the same firm. We find that: (i) inventors collect only 8% of the total private return from invention; (ii) entrepreneurs get over 44% of the total gains; (iii) bluecollar workers get about 26% of the gains and the rest goes to whitecollar workers. Moreover, entrepreneurs start with significant negative returns prior to the patent application, but their returns subsequently become highly positive.

## 1 Introduction

Innovation-based growth models feature entrepreneurs who decide R&D investments by maximizing expected innovation revenues net of innovation costs. While the firm-level effects of innovation have been extensively analyzed in the existing literature (e.g., see Griliches, 1990; Hall and Ziedonis, 2001; Blundell et al., 2002), the same is not true regarding the sharing of innovation-generated revenues. In particular, we do not have a good understanding of how innovation revenues are shared within firms, even though the innovation and the subsequent commercialization efforts are incurred not only by the inventor but also by her co-employees and by the owners in the firm.[1] An early important exception is Van Reenen (1996); while insightful, the data did not allow a closer look at who in the workforce of a firm benefits. This paper is a first attempt at filling this gap, as we merge individual income data, firm-level data, patenting data and IQ data in Finland over the period 1988-2012 to analyze the returns to invention for inventors and their co-workers or stakeholders within the same firm.

*Addresses- Aghion: College de France and London School of Economics (P.Aghion@lse.ac.uk). Akcigit: University of Chicago (uakcigit@uchicago.edu). Hyytinen: University of Jyväskylä (ari.t.hyytinen@jyu.fi). Toivanen: Aalto University School of Business and KU Leuven (otto.toivanen@aalto.fi). We thank Xavier Jaravel, Mark Schankerman, John Van Reenen and our discussants Pierre Azoulay and Heidi Williams for helpful comments. Hyytinen and Toivanen thank the Yrjö Jahnsson Foundation for financial support. The usual caveat applies.

[1]See Aghion and Tirole (1994).

Following Van Reenen (1996), most closely related to our analysis in this paper are Toivanen and Väänänen (2012), Bell et al. (2017), and Akcigit et al. (2017). Toivanen and Väänänen (2012) use Finnish patent and income data to study the return to inventors of US patents. They find strong and long-lasting impacts, especially for the inventors of highly cited patents. Bell et al. (2017) merge US individual fiscal data, test score information and US individual patenting data over the recent period to look at the lifecycle of inventors and the returns to invention. Akcigit et al. (2017) merge historical patent and individual census records to study, among other things, inventor compensation. We complement the existing literature by offering new evidence on the returns to inventors, but foremost by offering what to our knowledge is the first evidence on wage spillovers to non-innovating coworkers of different types.[2]

## 2 Data

Our data come from the following sources: First, the *Finnish longitudinal employer-employee data* (FLEED) which we exploit for the period 1988-2012. FLEED is an annual panel constructed from administrative registers of individuals, firms and establishments, maintained by Statistics Finland. It includes information on individuals' labor market status, salaries and other sources of income extracted from tax and other administrative registers. It also includes information on other individual characteristics, and employer and plant characteristics. Second, the *European Patent Office* data provide information on characteristics such as the inventor names and applicant names.[3] We have collected patent information on all patents with at least one inventor who registers Finland as his or her place of residence. We use data on all patents with a Finnish inventor up to and including 2012. Third, the *Finnish Defence Force* provides us with information on IQ test results for conscripts who did their military service in 1982 or later; all conscripts take the IQ test in the early stages of the service. These data contains the raw test scores of visuospatial, verbal and quantitative IQ tests. We follow Aghion et al. (2017) and use the visuospatial IQ percentiles.[4]

We limit our estimation sample to years 1994 – 2010 to allow for pre-trends in the early part of the data sample and to ensure sufficient coverage of patent applications in the late parts of the data. In this paper we focus attention on male workers who did their military service in 1982 or later (meaning they were born in 1961 or later). To ensure sufficient labor market participation (individuals enter FLEED at age 15), we require positive wage income in preceding 4 years of the included observations. Finally, we restrict attention to private sector employees because we can only identify coworkers in the private sector.

We identify an individual as a coworker or stakeholder within the same firm if he: 1) works in the inventing firm in the year of the patent application, and 2) is never an inventor himself. We study the following classes of coworkers or stakeholders within the same firm besides inventors:

[2] Kline et al. (2017) also study, using US data, the returns to invention for both the inventor and her coworkers.

[3] Here we want to thank the research project "Radical and Incremental Innovation in Industrial Renewal" by the VTT Research Centre (Hannes Toivanen, Olof Ejermo and Olavi Lehtoranta) for granting us access to the patent-inventor data they compiled.

[4] All the registry data is matched using individual identifiers. The matching of patent data to registry data is described in Aghion et al. (2017).

$(i)$ entrepreneurs,[5] $(ii)$ white-collar workers,[6] and $(iii)$ blue-collar workers.[7]

## 3 Regression equation

Our main regression equation takes the form:

$$\ln(wage_{itya}) = \alpha_i + \sum_{\tau=-4,..,10} \delta_\tau treated_i \times 1[t=\tau] + \sum_{\tau=-4,..,10} \alpha_\tau 1[t=\tau] + \sum_{y=1995,..,2012} \alpha_{year} 1[y=year]$$
$$+ \sum_{age=min(age)+2,..,max(age)} \alpha_{age} 1[a=age] + \varepsilon_{itya}, \tag{1}$$

where subscript $i$ denotes individual; subscript $y$ denotes calendar year ($y = 1995, ..., 2012$), $t$ denotes treatment time ($t = -4, ..., 10$), and $a$ denotes age in years ($a = min(age) + 2, ..., max(age)$).

Our specification includes: 1) individual fixed effects; 2) treatment time fixed effects, with $t = 0$ denoting the year of patent application (baseline is $t = -5$); calendar year fixed effects (baseline year 1994); and age fixed effects (baseline is $a \leq min(age) + 1$ which may vary across estimation samples). The variable $treated_i$ is an indicator variable taking value 1 if individual $i$ belongs to the treatment group (inventor or coworker of type $k$ = entrepreneur, blue-collar worker, white-collar worker) and 0 otherwise, and the $\alpha$'s denote the coefficients of the various fixed effects. We cluster standard errors at the individual level throughout.

We employ a conditional difference-in-difference approach whereby we first match each treated individual with a control individual.[8] The matching is done without replacement on an annual basis, starting from 1994. Due to the small number of potential control individuals, we use a 3-year period for entrepreneurs. We limit the potential control group to individuals who never invent and have never been coworkers of an inventor and who work in the private sector in the year of treatment. We use the following variables for matching: $(i)$ having at least an MSc; $(ii)$ having a STEM education; $(iii)$ working in manufacturing; $(iv)$ living in the South-West of Finland; $(v)$ age ($< 30$, $31-40$, $41-50$, $> 50$); $(vi)$ quintiles of the annual firm size distribution; and $(vii)$ having visuospatial IQ less than the 50th percentile, in the 51st – 80th, in the 81st – 90th, or above the 90th percentile.

We execute the matching separately for each treated group (inventor, entrepreneur, blue-collar worker, white-collar worker). This choice means that apart from inventors, the matching is done within the same socioeconomic group. For white-collar workers, we perform the matching separately within the following subcategories: $(i)$ senior managers, $(ii)$ senior workers, $(iii)$ junior managers, and $(iv)$ junior workers.[9]

[5] Individuals within the same firm are identified as entrepreneurs if: 1) they contribute to the entrepreneur pension system, and: 2) they own at least 50% of the company.

[6] These and the remaining individuals' job status are identified through the socio-economic status code contained in the FLEED.

[7] The merged data contain 15M observations on over 700K individuals who work in some 300K firms. 7033 individuals invent at least once (conditional on inventing, avg. #applications = 3.08, median = 1). The annual number of observations varies between 340K (in 1988) and 730K (from 2006 onwards). In the merged data, we have the following proportions of inventor and coworker observations: (1) inventors: 0.011; (2) entrepreneurs: 0.048; (3) white-collar workers: 0.270; (4) blue-collar workers: 0.316; (5) others: 0.355. See Table A1 in the online appendix for descriptive statistics on wage income.

[8] For a similar approach, see Jaravel et al. (2017). We implement one-to-one matching using the coarsened exact matching of Iacus et al. (2012).

[9] In this matching, "senior" and "junior" refer to socioeconomic status, not biological age.

# 4 Regression results

Table 1 shows the results for our baseline regression where we constrain the treatment effect $\delta_t$ to be constant both after the year of the patent application (i.e., $\delta_t = \delta_{post}$ for $t = 0, ..., 10$) and before that year (i.e., $\delta_t = \delta_{pre}$ for $t = -4, ..., -1$). In other words, we allow for constant but different post-treatment and pre-treatment (or anticipation) effects.

Table 1: Returns Estimation

| VARIABLES | **Inventor** | **Entrepreneur** | **Whitecollar** | **Bluecollar** |
|---|---|---|---|---|
| Treated × pre | 0.0417*** | -0.0153 | 0.00567 | -0.0107** |
| | (0.0133) | (0.0825) | (0.00402) | (0.00504) |
| Treated × post | 0.0511*** | 0.279*** | 0.0208*** | 0.0227*** |
| | (0.0162) | (0.0902) | (0.00463) | (0.00556) |
| Observations | 93,939 | 13,372 | 1,320,370 | 916,811 |
| R-squared | 0.329 | 0.180 | 0.347 | 0.256 |
| Number of individuals | 8,185 | 1,123 | 107,986 | 87,288 |
| Dependent variable | lnwage | lnwage | lnwage | lnwage |
| Age FE | YES | YES | YES | YES |
| Calendar year FE | YES | YES | YES | YES |
| Treatment year FE | YES | YES | YES | YES |
| Individual FE | YES | YES | YES | YES |
| Age × calendar year FE | NO | NO | NO | NO |
| Pre-treatment effects | YES | YES | YES | YES |
| Sample | Base | Base | Base | Base |

Notes: Standard errors in parentheses, and clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.1. Estimation samples are based on CEM one-to-one matching using annual data without replacement, starting from 1994 with the following matching criteria: 1) having a science education; 2) having at least an MSc; 3) working in manufacturing; 4) region (2 regions); 5) firm size (quintiles); and 6) visuospatial IQ (4 groups). For all groups but inventors, the matching is done within the socioeconomic group and for white-collar workers, withing sub-groups. The dependent variable is the natural log of the wage of the individual in a given year, measured in 2014 euros. Treated is an indicator variable that takes value one for each observation of an individual who belongs to the treatment group and is 0 otherwise, post is an indicator variable that takes value 1 in the year of receiving the treatment and thereafter and is 0 otherwise and pre is an indicator variable that takes value 1 in the last 4 years preceding the year of treatment and 0 otherwise. All specifications include a full set of calendar year dummies (base year 1994), age dummies (base age $\leq$ $min(age) + 1$), and a set of treatment time dummies for treatment years, $t = -4, ..., 10$ (base year$t = -5$). All specifications include the size of the firm (# employees) as a control variable and a dummy for missing employment information. The sample includes observations with treatment year $t = -5, ..., 10$.

We find that inventors earn on average a wage increase of 5% post invention, and earn on average 4% prior to invention starting 4 years before invention. This is similar in magnitude to what Toivanen and Väänänen (2012) report for annual returns a few years after the patent is granted. Next, we look at coworkers, and we find returns that are heterogenous across the different types of coworkers. Entrepreneurs earn the highest returns with almost 28% post invention, but nothing pre-invention. White collars earn a return of about 2% post invention but nothing pre-invention, and blue collars earn 2.3% post invention but lose 1% pre-invention. Incidentally, it is interesting to see that blue-collar workers experience a post-invention return on par or slightly higher than those experienced by white-collar workers.

We then turn to the full specification of equation 1. We display the results for inventors, entrepreneurs, bluecollar and whitecollar workers in Figure 1. Inventors earn returns already in anticipation of the patent application; after the patent application, there is a slight (though statistically indistinguishable) decrease, but soon, the returns start again to increase.

The estimated returns to entrepreneurs display a markedly different path. They start with significant negative returns in anticipation of the patent application but already before the year of patent application the returns turn positive. Then the entrepreneurs' returns keep rising and reach a maximum above 20% between two and two and half years after the invention time, with some fluctuations year to year thereafter. A potential explanation for the negative anticipation returns is that these entrepreneurs in innovative (and small) companies are credit constrained, and they finance invention partly by foregoing own consumption.

We have checked the robustness of these results in several ways. These results are reported in Table A2 of the online appendix for inventors and in Tables A3 – A5 for each of the three different types of coworkers. These robustness are the following: (*i*) excluding the anticipation effect (i.e., placing all observations with $t < 0$ into the base period). OLS estimates are shown in column 1 of each table, and fixed effects (FE) results in column 2. For comparison, we show OLS and FE results of our base specification (the latter is used in Table 1) in columns 3 and 4. (*ii*) We introduce the full set of age – calendar year interactions in columns 5 (OLS) and 6 (FE). (*iii*) We drop observations with missing information on the number of employees of the firm in columns 7 (OLS) and 8 (FE). (*iv*) We exclude observations from the top-3 employers of inventors in columns 9 and 10. (*v*) We use the log of the sum of wage and capital income as the dependent variable in columns 11 and 12. (*vi*) We include more base-period observations (i.e., observations with $t < -5$) in columns 13 and 14. (*vii*) As our last robustness test, we include observations where the individual works in the public sector (i.e., we don't observe a firm identifier; columns 15 and 16). Our results are robust to these changes with two expected exceptions: first, the estimated returns to inventors are reduced when we don't allow for anticipation effects (Table A1, column 2) which were estimated to be positive (Table 1). Second, the estimated returns to entrepreneurs are lower (0.13) and not statistically significant if we exclude observations with missing information on the number of employees (Table A3, column 8). With this rule, we lose 20% of the estimation sample of entrepreneurs as the rule excludes mainly observations from small, often entrepreneur-driven, firms.

An important aspect of the returns to invention is an understanding of how the proceeds from invention are shared among the different types of workers within the innovating firm. To illustrate this, we use the coworker-type specific return estimates from Figure 1, the shares of different types of coworkers in innovating firms (we use 2003 data) and the wages of different types of coworkers in innovating firms before invention (we use mean wages in our base year, i.e., $t = -5$). Using these numbers, we calculated both the total dollar-increase in the wage bill of an innovating firm, and how it is shared between these different types of workers. The result, displayed in Figure 2, reveals some interesting conclusions: First, inventors get only 8% of the total gains; second, entrepreneurs get over 44% of the total gains; and finally, bluecollar workers get about 26% of the gains and the rest goes to the whitecollar workers.

## 5 Conclusion

In this paper we start closing the gap on providing evidence on income spillovers from invention within the inventing firm. Using data from Finland 1988 - 2012 we found significant returns to inventors themselves. Moreover, we found significant spillover effects within the firm, with non-inventing coworkers and entrepreneurs in the same firm also benefitting from the invention. Both white-collar and blue-collar workers benefit from invention; after the invention, if anything, the latter more than the former. Entrepreneurs experience the highest percentage annual gains at over 20%. Gains for all groups are long-lasting.

Our findings show that inventors collect only less than 10% of the total private return. This result highlights the importance of taking into account the incentives of other actors in the firm (e.g., firm owner and co-workers) who also benefit from an invention both in modeling invention and in drawing policy conclusions (e.g., on taxation).

## References

Aghion, P., U. Akcigit, A. Hyytinen, and O. Toivanen: 2017, 'The Social Origins of Inventors'. National Bureau of Economic Research WP 24110.

Aghion, P. and J. Tirole: 1994, 'The Management of Innovation'. *Quarterly Journal of Economics* **109**(4), 1185–1209.

Akcigit, U., J. Grigsby, and T. Nicholas: 2017, 'The Rise of American Ingenuity: Innovation and Inventors of the Golden Age'. National Bureau of Economic Research WP 23047.

Bell, A., R. Chetty, X. Jaravel, N. Petkova, and J. Van Reenen: 2017, 'Who Becomes an Inventor in America? The Importance of Exposure to Innovation'. National Bureau of Economic Research WP 24062.

Blundell, R., R. Griffith, and F. Windmeijer: 2002, 'Individual Effects and Dynamics in Count Data Models'. *Journal of Econometrics* **108**(1), 113–131.

Griliches, Z.: 1990, 'Patent Statistics as Economic Indicators: A Survey'. *Journal of Economic Literature* **28**(4), 1661–1707.

Hall, B. H. and R. H. Ziedonis: 2001, 'The Patent Paradox Revisited: An Empirical Study of Patenting in the US Semiconductor Industry, 1979–1995'. *RAND Journal of Economics* **32**(1), 101–128.

Iacus, S. M., G. King, and G. Porro: 2012, 'Causal Inference without Balance Checking: Coarsened Exact Matching'. *Political Analysis* **20**(1), 1–24.

Jaravel, X., N. Petkova, and A. Bell: 2017, 'Team-specific Capital and Innovation'. *American Economic Review*. forthcoming.

Kline, P., N. Petkova, H. Williams, and O. Zidar: 2017, 'Who Profits from Patents? Rent-Sharing at Innovative Firms'. Working Paper.

Toivanen, O. and L. Väänänen: 2012, 'Returns to Inventors'. *Review of Economics and Statistics* **94**(4), 1173–1190.

Van Reenen, J.: 1996, 'The Creation and Capture of Rents: Wages and Innovation in a Panel of U.K. Companies'. *Quarterly Journal of Economics* **111**(1), 195–226.

Figure 1: Returns Estimations

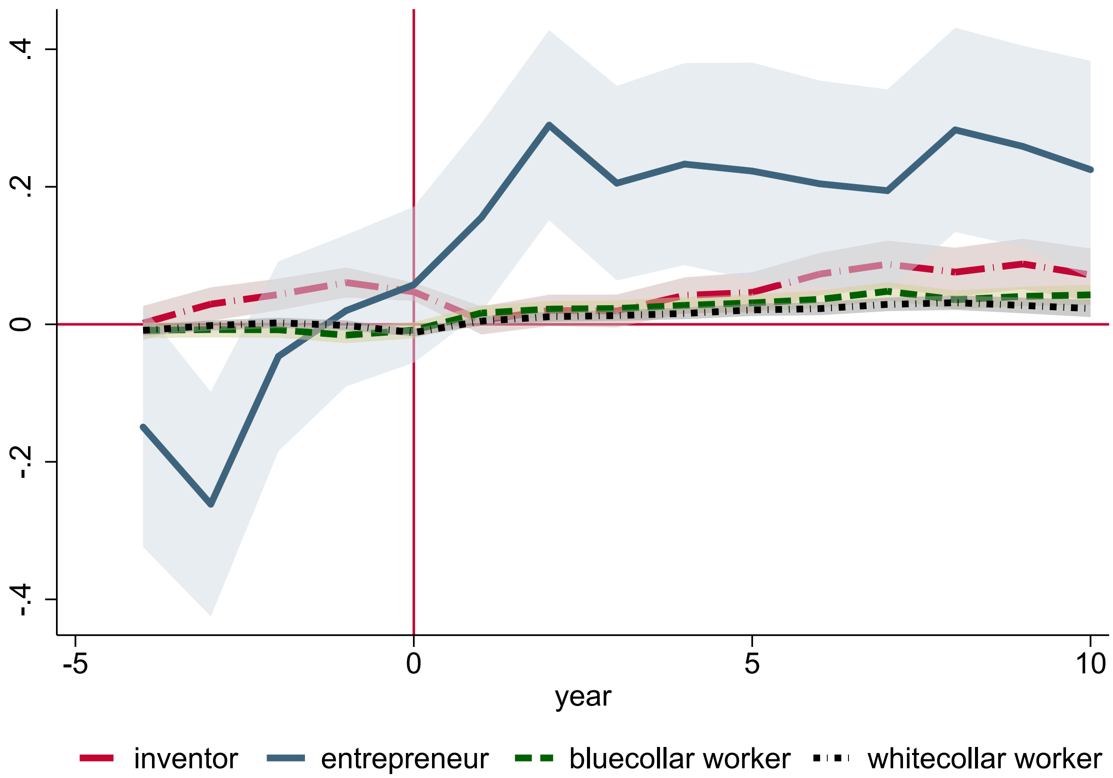
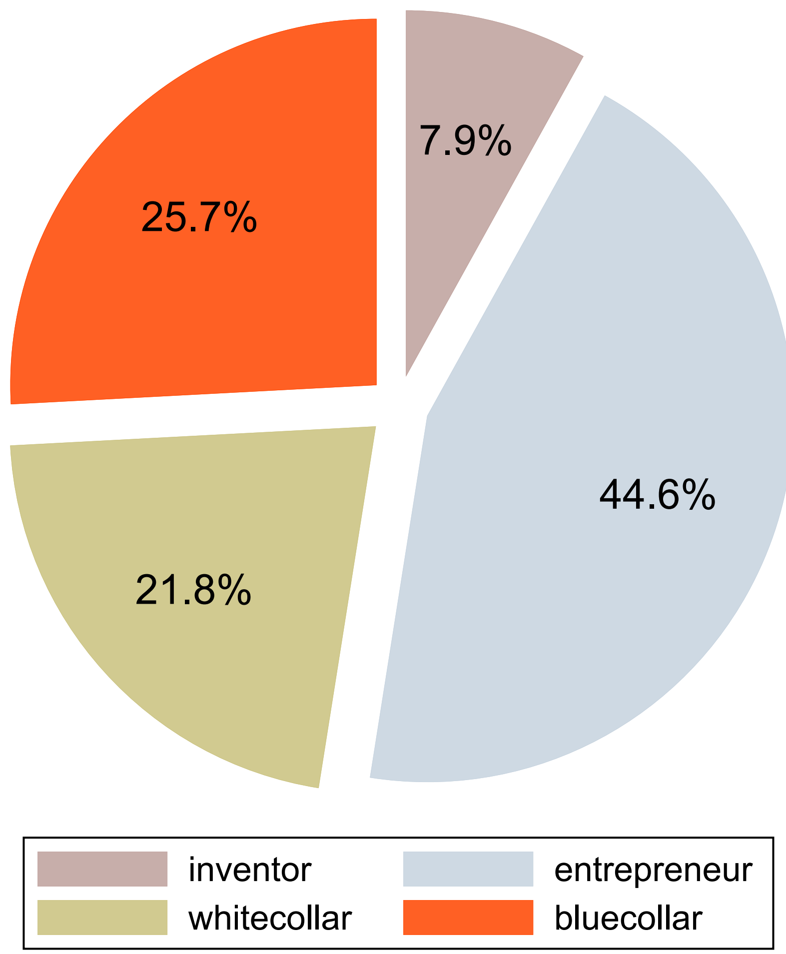


Figure 2: Returns Distribution



# Online Appendix for "On the Returns to Invention Within Firms: Evidence from Finland"

Philippe Aghion  Ufuk Akcigit  Ari Hyytinen  Otto Toivanen

January 14, 2018

Table A1: Descriptive Statistics

*– Treatment group –*

| | | Inventor | Entrepreneur | Whitecollar | Bluecollar |
|---|---|---|---|---|---|
| Wage before | mean | 42,938 | 28,278 | 37,231 | 27,828 |
| | sd | 41,220 | 26,366 | 30,742 | 12,116 |
| | median | 39,642 | 21,507 | 34,942 | 28,173 |
| Wage pre | mean | 48,536 | 29,946 | 41,213 | 29,067 |
| | sd | 33,212 | 25,834 | 28,158 | 12,611 |
| | median | 45,169 | 24,819 | 38,078 | 29,296 |
| Wage post | mean | 66,788 | 47,134 | 51,414 | 35,639 |
| | sd | 67,992 | 29,803 | 35,614 | 14,037 |
| | median | 60,163 | 41,636 | 46,092 | 34,914 |

*– Control group –*

| | | Inventor | Entrepreneur | Whitecollar | Bluecollar |
|---|---|---|---|---|---|
| Wage before | mean | 38,799 | 26,377 | 34,690 | 26,328 |
| | sd | 27,910 | 25,166 | 22,090 | 11,460 |
| | median | 35,652 | 21,683 | 32,891 | 26,747 |
| Wage pre | mean | 43,361 | 28,239 | 38,773 | 27,474 |
| | sd | 30,736 | 24,396 | 32,343 | 11,783 |
| | median | 39,335 | 22,717 | 35,982 | 27,789 |
| Wage post | mean | 55,600 | 39,779 | 49,035 | 32,935 |
| | sd | 34,413 | 32,980 | 29,959 | 12,827 |
| | median | 48,852 | 32,849 | 44,040 | 32,476 |

Notes: the wages are in real year 2014 euros. "Before" refers to years five or more years before the patent application, "pre" to years 1-4 years before the patent application, and "post" to the year of the patent application, and the following 10 years.

Table A2: Robustness Results for Inventors

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) | (14) | (15) | (16) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Treated | 0.131*** | | 0.101*** | | 0.102*** | | 0.120*** | | 0.0920*** | | 0.115*** | | 0.102*** | | 0.0902*** | |
| | (0.0105) | | (0.0163) | | (0.0161) | | (0.0177) | | (0.0170) | | (0.0164) | | (0.0145) | | (0.0180) | |
| Treated pre | | | 0.0360*** | 0.0417*** | 0.0337** | 0.0398*** | 0.0362** | 0.0396*** | 0.0418*** | 0.0453*** | 0.0356** | 0.0431*** | 0.0345*** | 0.0394*** | 0.0573*** | 0.0602*** |
| | | | (0.0139) | (0.0133) | (0.0138) | (0.0132) | (0.0152) | (0.0143) | (0.0147) | (0.0140) | (0.0139) | (0.0131) | (0.0124) | (0.0121) | (0.0161) | (0.0156) |
| Treated post | 0.0187* | 0.0159* | 0.0488*** | 0.0511*** | 0.0434*** | 0.0468*** | 0.0621*** | 0.0697*** | 0.0544*** | 0.0508*** | 0.0487*** | 0.0548*** | 0.0469*** | 0.0493*** | 0.0597*** | 0.0605*** |
| | (0.0102) | (0.00966) | (0.0167) | (0.0162) | (0.0165) | (0.0161) | (0.0182) | (0.0171) | (0.0175) | (0.0169) | (0.0167) | (0.0158) | (0.0152) | (0.0149) | (0.0185) | (0.0186) |
| Observations | 93,939 | 93,939 | 93,939 | 93,939 | 93,939 | 93,939 | 76,717 | 76,717 | 87,097 | 87,097 | 94,006 | 94,006 | 98,527 | 98,527 | 97,013 | 97,013 |
| R-squared | 0.271 | 0.329 | 0.271 | 0.329 | 0.282 | 0.337 | 0.260 | 0.298 | 0.259 | 0.322 | 0.270 | 0.329 | 0.291 | 0.357 | 0.382 | 0.433 |
| Number of individuals | | 8,185 | | 8,185 | | 8,185 | | 7,439 | | 7,982 | | 8,185 | | 8,185 | | 8,185 |
| Dependent variable | lnwage | lnwage | lnwage | lnwage | lnwage | lnwage | lnwage | lnwage | lnwage | lnwage | lnearnings | lnearnings | lnwage | lnwage | lnwage | lnwage |
| Individual FE | NO | YES | NO | YES | NO | YES | NO | YES | NO | YES | NO | YES | NO | YES | NO | YES |
| Age FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Calendar year FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Treatment year FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Age x calendar year FE | NO | NO | NO | NO | YES | YES | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| Pre-treatment effects | NO | NO | YES | YES | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| Sample | A | A | A | A | A | A | B | B | C | C | A | A | D | D | E | E |

*Notes*: Standard errors in parentheses, and clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.1. All specifications include the size of the firm (# employees) as a control variable and a dummy for missing employment information. Sample details are as follows. A: Base, B: Drop observations with missing information on number of employees of the firm, C: exclude top-3 employers of inventors, D: include observations for which $t < -5$, E: include observations where individual in the public sector.

Table A3: Robustness Results for Entrepreneurs

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) | (14) | (15) | (16) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Treated | 0.0808 | | 0.148 | | 0.153 | | 0.220* | | 0.148 | | 0.0995 | | 0.160* | | 0.0242 | |
| | (0.0579) | | (0.0941) | | (0.0974) | | (0.120) | | (0.0941) | | (0.0884) | | (0.0847) | | (0.0842) | |
| Treated pre | | | -0.0792 | -0.0153 | -0.0915 | -0.0226 | -0.152 | -0.0822 | -0.0792 | -0.0153 | -0.0637 | 0.00974 | -0.0899 | -0.0604 | -0.0302 | -0.00825 |
| | | | (0.0856) | (0.0825) | (0.0883) | (0.0854) | (0.116) | (0.106) | (0.0856) | (0.0825) | (0.0851) | (0.0831) | (0.0757) | (0.0730) | (0.0766) | (0.0750) |
| Treated post | 0.213*** | 0.292*** | 0.146 | 0.279*** | 0.142 | 0.264*** | 0.0394 | 0.126 | 0.146 | 0.279*** | 0.172* | 0.297*** | 0.134 | 0.232*** | 0.263*** | 0.340*** |
| | (0.0523) | (0.0512) | (0.0937) | (0.0902) | (0.0971) | (0.0930) | (0.124) | (0.113) | (0.0937) | (0.0902) | (0.0896) | (0.0883) | (0.0825) | (0.0800) | (0.0863) | (0.0835) |
| Observations | 13,372 | 13,372 | 13,372 | 13,372 | 13,372 | 13,372 | 10,707 | 10,707 | 13,372 | 13,372 | 13,455 | 13,455 | 13,678 | 13,678 | 15,371 | 15,371 |
| R-squared | 0.129 | 0.180 | 0.129 | 0.180 | 0.166 | 0.219 | 0.103 | 0.134 | 0.129 | 0.180 | 0.118 | 0.161 | 0.133 | 0.186 | 0.186 | 0.258 |
| Number of individuals | | 1,123 | | 1,123 | | 1,123 | | 1,087 | | 1,123 | | 1,124 | | 1,123 | | 1,123 |
| Dependent variable | lnwage | lnwage | lnwage | lnwage | lnwage | lnwage | lnwage | lnwage | lnwage | lnwage | lnearnings | lnearnings | lnwage | lnwage | lnwage | lnwage |
| Individual FE | NO | YES | NO | YES | NO | YES | NO | YES | NO | YES | NO | YES | NO | YES | NO | YES |
| Age FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Calendar year FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Treatment year FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Age x calendar year FE | NO | NO | NO | NO | YES | YES | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| Pre-treatment effects | NO | NO | YES | YES | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| Sample | A | A | A | A | A | A | B | B | C | C | A | A | D | D | E | E |

*Notes*: Standard errors in parentheses, and clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.1. All specifications include the size of the firm (# employees) as a control variable and a dummy for missing employment information. Sample details are as follows. A: Base, B: Drop observations with missing information on number of employees of the firm, C: exclude top-3 employers of inventors, D: include observations for which $t < -5$, E: include observations where individual in the public sector.

## Table A4: Robustness Results for White-collar workers

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) | (14) | (15) | (16) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Treated | 0.0211*** | | 0.0259*** | | 0.0254*** | | 0.0515*** | | 0.0259*** | | 0.0272*** | | 0.0252*** | | 0.0183*** | |
| | (0.00321) | | (0.00492) | | (0.00490) | | (0.00581) | | (0.00492) | | (0.00495) | | (0.00437) | | (0.00539) | |
| Treated pre | | | -0.00579 | 0.00567 | -0.00568 | 0.00580 | 0.00774 | 0.00805* | -0.00579 | 0.00567 | -0.00830* | 0.00396 | -0.00533 | 0.00706* | -0.00235 | 0.00229 |
| | | | (0.00429) | (0.00402) | (0.00428) | (0.00401) | (0.00513) | (0.00458) | (0.00429) | (0.00402) | (0.00427) | (0.00399) | (0.00377) | (0.00360) | (0.00486) | (0.00466) |
| Treated post | -0.000719 | 0.0160*** | -0.00553 | 0.0208*** | -0.00552 | 0.0210*** | 0.0226*** | 0.0235*** | -0.00553 | 0.0208*** | -0.00793 | 0.0178*** | -0.00489 | 0.0223*** | -0.000632 | 0.0184*** |
| | (0.00295) | (0.00276) | (0.00488) | (0.00463) | (0.00487) | (0.00463) | (0.00587) | (0.00537) | (0.00488) | (0.00463) | (0.00489) | (0.00460) | (0.00435) | (0.00421) | (0.00539) | (0.00529) |
| Observations | 1,320,370 | 1,320,370 | 1,320,370 | 1,320,370 | 1,320,370 | 1,320,370 | 872,101 | 872,101 | 1,320,370 | 1,320,370 | 1,320,676 | 1,320,676 | 1,390,720 | 1,390,720 | 1,362,446 | 1,362,446 |
| R-squared | 0.320 | 0.347 | 0.320 | 0.347 | 0.324 | 0.348 | 0.320 | 0.329 | 0.320 | 0.347 | 0.320 | 0.348 | 0.336 | 0.373 | 0.421 | 0.456 |
| Number of individuals | | 107,986 | | 107,986 | | 107,986 | | 87,224 | | 107,986 | | 107,986 | | 107,986 | | 107,986 |
| Dependent variable | lnwage | lnwage | lnwage | lnwage | lnwage | lnwage | lnwage | lnwage | lnwage | lnwage | lnearnings | lnearnings | lnwage | lnwage | lnwage | lnwage |
| Individual FE | NO | YES | NO | YES | NO | YES | NO | YES | NO | YES | NO | YES | NO | YES | NO | YES |
| Age FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Calendar year FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Treatment year FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Age x calendar year FE | NO | NO | NO | NO | YES | YES | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| Pre-treatment effects | NO | NO | YES | YES | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| Sample | A | A | A | A | A | A | B | B | C | C | A | A | D | D | E | E |

*Notes*: Standard errors in parentheses, and clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.1. All specifications include the size of the firm (# employees) as a control variable and a dummy for missing employment information. Sample details are as follows. A: Base, B: Drop observations with missing information on number of employees of the firm, C: exclude top-3 employers of inventors, D: include observations for which $t < -5$, E: include observations where individual in the public sector.

### Table A5: Robustness Results for Blue-collar workers

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) | (14) | (15) | (16) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Treated | 0.0552*** | | 0.0657*** | | 0.0661*** | | 0.0589*** | | 0.0657*** | | 0.0661*** | | 0.0641*** | | 0.0503*** | |
| | (0.00328) | | (0.00582) | | (0.00581) | | (0.00614) | | (0.00582) | | (0.00586) | | (0.00490) | | (0.00672) | |
| Treated pre | | | -0.0124** | -0.0107** | -0.0122** | -0.0107** | -0.0126** | -0.00878* | -0.0124** | -0.0107** | -0.0130** | -0.0115** | -0.0111** | -0.00785* | -0.00814 | -0.00461 |
| | | | (0.00550) | (0.00504) | (0.00549) | (0.00503) | (0.00583) | (0.00527) | (0.00550) | (0.00504) | (0.00552) | (0.00506) | (0.00461) | (0.00431) | (0.00641) | (0.00611) |
| Teated post | 0.0332*** | 0.0319*** | 0.0226*** | 0.0227*** | 0.0218*** | 0.0220*** | 0.0307*** | 0.0246*** | 0.0226*** | 0.0227*** | 0.0203*** | 0.0198*** | 0.0240*** | 0.0257*** | 0.0332*** | 0.0374*** |
| | (0.00319) | (0.00318) | (0.00585) | (0.00556) | (0.00584) | (0.00555) | (0.00618) | (0.00580) | (0.00585) | (0.00556) | (0.00589) | (0.00558) | (0.00496) | (0.00486) | (0.00671) | (0.00664) |
| Observations | 916,811 | 916,811 | 916,811 | 916,811 | 916,811 | 916,811 | 783,114 | 783,114 | 916,811 | 916,811 | 917,056 | 917,056 | 950,349 | 950,349 | 948,915 | 948,915 |
| R-squared | 0.229 | 0.256 | 0.229 | 0.256 | 0.231 | 0.259 | 0.228 | 0.258 | 0.229 | 0.256 | 0.230 | 0.257 | 0.235 | 0.269 | 0.339 | 0.367 |
| Number of individuals | | 87,288 | | 87,288 | | 87,288 | | 82,582 | | 87,288 | | 87,288 | | 87,288 | | 87,288 |
| Dependent variable | lnwage | lnwage | lnwage | lnwage | lnwage | lnwage | lnwage | lnwage | lnwage | lnwage | lnearnings | lnearnings | lnwage | lnwage | lnwage | lnwage |
| Individual FE | NO | YES | NO | YES | NO | YES | NO | YES | NO | YES | NO | YES | NO | YES | NO | YES |
| Age FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Calendar year FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Treatment year FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Age x calendar year FE | NO | NO | NO | NO | YES | YES | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| Pre-treatment effects | NO | NO | YES | YES | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| Sample | A | A | A | A | A | A | B | B | C | C | A | A | D | D | E | E |

*Notes*: Standard errors in parentheses, and clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.1. All specifications include the size of the firm (# employees) as a control variable and a dummy for missing employment information. Sample details are as follows. A: Base, B: Drop observations with missing information on number of employees of the firm, C: exclude top-3 employers of inventors, D: include observations for which $t < -5$, E: include observations where individual in the public sector.