Paper to be presented at the

DRUID Society Conference 2014, CBS, Copenhagen, June 16-18

# How much does it cost to be a scientist?

**Benjamin Balsmeier**
KU Leuven
Department of Managerial Economics, Strategy and Innovation
benjamin.balsmeier@kuleuven.be

**Maikel Pellens**
KULeuven
Department of Managerial Economics, Strategy, and Innovation
maikel.pellens@kuleuven.be

Abstract

This paper examines the academe-industry wage gap. We show that self-selection in the scientific labor market inflates the wage gap by approximately 20%. Differences with regard to research and development activities explain a significant part of the wage differential. Academic scientists who spend much time on research face much lower counterfactual wages in comparable positions in industry, whereas academic scientists who spend much time on development activities face much higher ones. This finding challenges the idea of a solely negative relationship between science and wages. Our investigation further shows that preferences for science moderate the relationship between research orientation and wages. The results have important implications for managers that try to hire academic scientists, policy makers that aim to increase development oriented research activities at universities, and individual scientists thinking about whether to pursue a career in industry or academe.

Jelcodes:O31,O38

# How Much Does It Cost to Be a Scientist?

**Abstract**
This paper examines the academe-industry wage gap. We show that self-selection in the scientific labor market inflates the wage gap by approximately 20%. Differences with regard to research and development activities explain a significant part of the wage differential. Academic scientists who spend much time on research face much lower counterfactual wages in comparable positions in industry, whereas academic scientists who spend much time on development activities face much higher ones. This finding challenges the idea of a solely negative relationship between science and wages. Our investigation further shows that preferences for science moderate the relationship between research orientation and wages. The results have important implications for managers that try to hire academic scientists, policy makers that aim to increase development oriented research activities at universities, and individual scientists thinking about whether to pursue a career in industry or academe.

**JEL-Classification:** G24, O31, 038
**Keywords:** Academe-Industry Wage Gap, Economics of Science, Matching, Taste for Science

# 1. Introduction

The National Science Foundation (NSF) Science and Engineering indicators 2012 report that the median wage of a doctorate recipient up to five years after graduation is $85,000 in industry, compared to $65,000 in academe (National Science Board, 2012). This difference is one of the main reasons that lead academic scientists to seek a position in industry (Walker, Vignoles, and Collins, 2010). Although monetary rewards represent only one component of the scientific reward system, they are still highly valued by individual researchers. In a recent survey of more than 5,000 life scientists and physical scientists 37 percent of the academic scientists and 47 percent of the industrial scientists rated salary as a "very important" job factor (see Sauermann and Stephan, 2012, p. 9).

Despite the fact that the academe-industry wage differential plays a major role in scientists' career decisions and, consequently, the production of knowledge in an economy, we know relatively little about the processes underlying the wage gap and how it may vary across individual scientists. Up to now the debate centers around earnings profiles within each sector and the scientific orientation of industrial research.[1]

A sparse literature tries to estimate the academe-industry wage difference. Agarwal and Ohyama (2013), for instance, compare wages of basic and applied scientists in industry and academe. While basic scientists earn less in academe than applied scientists, higher complementarities between applied and basic science lead to smaller wage differences within

[1] See for instance Stern (2004) or Sauermann & Roach (2013) for industrial wage dynamics. Many literature studies academic wage dynamics, including factors such as gender (Barbezat, 1987; Bayer and Astin, 1968; McNabb and Wass, 1997), seniority (Moore, Newman, and Turnball, 1998), and international wage differences (Ong and Mitchell, 2000; Altbach, Reisberg, Yudkevich, Androushcak, and Pacheco, 2012; Stevens, 2004). An earlier stream of research estimates academics' earnings functions, finding that scientists' earnings are concave, peaking late in their career (Stephan, 1996; Diamond, 1986a; Laitner and Stafford, 1985; Lillard and Weiss, 1979; Weiss and Lillard, 1978; Creedy, 1988).

industry. It remains an open question though, how much of the academe-industry wage gap is driven by known self-selection in the scientific labor market and how much of it is influenced by researchers' focus on basic research as opposed to development oriented research. Some other studies examine the wage gap by comparing wages of academics with other workers' wages (Walker et al., 2010; Stevens, 2004). While these studies provide useful indications of the wage differences between academe and other sectors, they are limited in the sense that they compare wages of academics with those of non-academics (or, as in Walker et al. (2010), a selected subsample thereof).

We advance the literature by estimating the academe-industry wage gap controlling for known sources of self-selection and differences in research focus between industry and academe. A rich dataset of Belgian Ph.D. holders allows us to match to each academic scientist an industrial scientist that is comparable in preferences, time spent on research, time spent on developing new products, as well as a variety of personal characteristics. Observing the time spent on research and development crucially allow us to account for differences in research activities in a highly detailed manner, whereas other studies (e.g. Agarwal and Ohyama, 2013) have relied on a categorical classification of basic and applied research. Once we correct for self-selection and differences in research activities, the average wage gap drops from 28.3% to 22.5%. Our analysis further uncovers a remarkable heterogeneity in the wage gap: an academic who spends much time on research faces a significantly lower wage gap, while the wage gap is much higher for academics who spend much time on development.

The analysis also complements Stern's (2004) and Sauermann and Roach's (2013) examinations of the science-wage relationship by showing that development activities relate positively to the wage gap. This is enabled by the detailed measurement of time spent on research and development, whereas Stern (2004) uses relatively broad categories, i.e. the chance to continue previous research and the permission to publish, in order to quantify a job's scientific environment. Sauermann and Roach (2013) focus on a scientist's reservation wage to explain the acceptance of a job without the opportunity to publish versus a job with that opportunity. We finally show that preferences have a moderating role: scientists with a high taste for science trade off monetary rewards against research orientation at a much higher level than those scientists with a rather low taste for science.

Overall, these findings have important implications for career decisions of scientists, managers trying to hire academic scientists, and policy makers aiming to enhance development oriented research activities at universities. A typical academic scientist, who has a strong taste for science and who spends a large portion of his time conducting pure research activities, might earn significantly less in industry than expected from a naïve average wage comparison. Moreover, the wage gap is by far the largest for academic researchers working on product developments. As academics are more and more urged to engage in development orientated research projects, careers in industry become more attractive, which in turn increases the risk of an academic brain drain. Scientists with a high taste for science, however, might still be better off in academe. Lastly, the significant remaining wage gap after matching indicates that factors other than preferences and research activities, e.g. differences in research valorization, are responsible for a large part of the academe-industry wage differential.

## 2. Conceptual framework

The conceptual framework focuses on two mechanisms that affect the academe-industry wage gap: a preference effect stemming from the willingness of scientists to forgo monetary rewards in return for a stronger research orientation, and a potential productivity effect stemming from shared returns to research between scientists and the institutions they work for. We start with a brief description of differences in the scientific reward systems in academe and industry.

### 2.1. The scientific reward system in academe and industry

Wage differences between academe and industry are arguably driven by differences in the organization of research in both sectors (e.g. Aghion, Dewatripont, and Stein, 2008; Sauermann and Stephan, 2012). Especially in European countries, where academics at higher education institutions are hired as civil cervants, earnings profiles are remarkably flat. Wages in academe are further not strongly dependent on scientific productivity. Monetary incentive schemes are usually low powerd or non-existent because of the difficulty to monitor academic research and the risk of encouraging scientists to focus on safe projects with guaranteed results (Konrad and Pfeffer, 1990). Some academics increase their income with consulting, speaking fees, or prizes. Other components of the scientific reward system, including the intellectual challenge involved with new projects, pleasure from "puzzle solving", and peer recognition by other scholars, are nonmonetary in nature (Aghion, Dewatripont, and Stein, 2008; Merton, 1973; Dasgupta and David, 1994; Stephan, 1996). Academic institutions generally rely on the nonmonetary reward system to incentivize their researchers to engage in the most fruitful research projects and knowledge production.

Industrial companies, on the other hand, are supposed to engage in science to maximize profits, focusing on applied research topics with high expected returns from new product developments (Rosenberg, 1990; Cohen and Levinthal, 1989, 1990; Arora and Gambardella, 1994). Industrial researchers thus tend to work on more narrowly defined research projects with emphasis on the potential commercialization of the output.

Stern (2004) shows that this has different implications for the relationship between science and wages in industry. On the one hand, researchers in industry could be paid a compensating differential for foregone freedom in choosing the kind of research they prefer to follow (the preference hypothesis). This leads to a negative relation between research orientation and wages in industry (similar results have been found by Aghion, Dewatripont, and Stein, 2008; Agarwal and Ohyama, 2013; Sauermann and Roach, 2013; Sauermann and Stephan, 2012; Fini and Lacetera, 2010; Lacetera, 2009). On the other hand, firms might participate in science to increase their absorptive capacity, capture spillovers and raise their R&D productivity (the productivity hypothesis). If firms share at least a part of those productivity gains with their scientists, we might expect a positive relation between research orientation and wages in industry (see also Cohen and Levinthal, 1990; Rosenberg, 1990). Depending on which effect dominates, this has important implications for the size of the academe-industry wage gap, and how it varies across different profiles of individual scientists.

Stern's empirical analysis of young biologists' job offers revealed a negative relation between the scientific orientation of a particular job and the offered wage. He concluded that the preference effect dominates potential productivity effects. In line with this finding, Sauermann and Roach (2013) recently showed that scientists with a high taste for science demand higher salaries if they have to give up the right to publish research findings. These findings do not indicate that there are no productivity effects at all, but rather that the preference effect on average dominates potential productivity effects.

Extant studies employ quite broad proxies of 'science': Stern (2004) uses continuation of research and permission or incentives to publish in order to quantify a job's scientific environment, and Sauermann and Roach (2013) use the ability to publish. This focus on specific aspects of 'science' neglects that a scientist's job in practice will be somewhere in a spectrum ranging from completely basic to completely applied. Moreover, both of the latter studies rely on supply-side dynamics identified through surveys of scientists, instead of equilibria observed in actual job outcomes. By analyzing the time scientists actually spend on 'research' and 'development' in their jobs, we hope to generate new insights concerning the preference and productivity effects underlying scientists' wages in industry, and hence the wage gap.

### 2.2. Wage effects of time spent on research versus development

In this section, we derive hypotheses on the size of the wage gap as a function of a stronger focus on research or development. Specifically, we distinguish between the proportion of time a scientist spends on research activities which are not directly related to product or process development and the proportion of time a scientist spends on the development of new products and processes. We argue that both activities differ in how they are valued by individual scientists as well as the value they generate for the firm, and, hence, whether a preference or productivity effect dominates. While Stern (2004) used permission to publish research findings, permission to continue prior research and monetary incentives to publish as indicators for scientific orientation (p. 843), our measure explicitly captures the applied or basic focus of the researchers' work. This allows us to disentangle – conceptually and empirically – a domination of productivity and preference effects related to basic as opposed to applied research. Thereby, the analysis sheds new light on the question to which extent preference effects and productivity effects of science apply, and how these vary with basic versus applied research activities.[2]

[2] Note that while we make assumptions about the relative productivity of research and development, these assumptions are not essential for the analysis. The main purpose is to show that research and development can be expected to have different productivity effects, and that it is important to account for these. In practice, research and development can be complementary, and whether research or development is more productive depends on their relative marginal productivity.

As ‘research’ comes closest to Sterns’ (2004) definition of science (as compared to development), we expect the preference effect to dominate the productivity effect. The underlying assumption is that spending time on research satisfies researchers’ preferences for science, leading them to accept lower wages in industry. Firms might also raise their productivity through research, but the shared productivity gains are likely smaller than the voluntarily forgone rewards (Stern, 2004).

Since development in industry is largely determined by strategic considerations and the production and commercialization process (Karlsson et al., 2004), it offers limited freedom to individual researchers and is therefore less satisfying than research. Hence, we expect scientists to be less likely to trade off wage against time spend on development, possibly leading to a positive preference effect where firms pay a compensating differential for foregone freedom. At the same time, product and process development is supposed to be less risky and to generate more revenue in the short term than research. Further, since it is less risky, firms are better able to measure successes, allowing for monetary incentive schemes by which scientists participate in the returns to their research activities. For these reasons, we expect industrial scientists’ wages to be positively affected by the proportion of time spent on development oriented research. Figure 1 provides a corresponding illustration of how we expect preference and productivity effects of research and development to affect wages in industry. Since preference effects dominate in research and productivity effects dominate in development, industrial researchers’ wages decrease with time spent on research and increase with time spent on development.

**Figure 1: Productivity versus preference effects of research and development**

| **Type of R&D activity** | Research | | | Development | | |
|---|---|---|---|---|---|---|
| | preference effect | productivity effect | **net impact** | preference effect | productivity effect | **net impact** |
| **Effect on wage** | ↓ | ↑ | ↓ | ↓ | ↑ | ↑ |

In academe, spending more time on research does not affect scientists' wages in a direct fashion,[3] while valorizing research results through development might generate additional revenue streams. However, it is difficult to assess how much academic research would profit from development compared to their counterparts in industry. They may profit more, since they do not have to share returns with their employers. They might also profit less, since universities have less assets, knowledge, and incentives to successfully commercialize than firms. Combined with our earlier assessment of industry wage dynamics, this leads us to formulate the following hypotheses regarding the effects of different research activities on the wage gap.

**Hypothesis 1:** The wage gap an academic scientist faces decreases with the proportion of time spent on research.

**Hypothesis 2a:** The wage gap an academic scientist faces increases with the proportion of time spent on development.

**Hypothesis 2b:** The wage gap an academic scientist faces decreases with the proportion of time spent on development.

[3] However, research success could increase one's chances for promotion, and hence increase earnings in the long run.

All hypotheses have direct implications for the average academe-industry wage differential, but the size depends on the marginal impacts of spending more time on research and development on wages, and the factual difference between the average time scientists spend on research and development in academe compared to industry. Therefore, we refrain from formulating hypotheses about whether the wage gap will become wider or narrower once these mechanisms have been accounted for. A detailed evaluation follows in section 5.

### 2.3. Scientists' preferences

Apart from differences in research activities between academe and industry the preferences of researchers and scientists might explain a large part of the academe-industry wage differential. The reasons are twofold. First, preferences determine how much compensation an industrial scientist demands for more or less scientific orientation of his job. Second, scientists select themselves into one of the two sectors according to their preferences.

A well-established literature has shown that scientists' choices for a career in academe or industry are largely affected by their 'taste for science' (Roach and Sauermann, 2010; Sauermann and Roach, 2012; Stern 2004). Researchers who have a high taste for science are considered to have a high intrinsic motivation to perform research, a desire for intellectual challenge, and strong preferences for classic aspects of the academic system, e.g. freedom in choosing research topics, focus on basic research questions, and rewards through peer recognition (Merton, 1973; Stephan and Levin, 1992; Roach and Sauermann, 2010; Stern, 2004; Sauermann and Roach, 2013).[4] As taste for science rises, scientists derive higher utility from

[4] The scientists' taste for commercialization, on the other hand, represents motivation through extrinsic factors, such as wages, extralegal benefits, and job security. Noteworthy, both preferences are not mutually exclusive, meaning that scientists can have a high taste for science as well as a high taste for commercialization (Roach & Sauermann, 2010; Sauermann & Roach, 2012; Agarwal & Ohyama, 2012). We include the researchers' taste for commercialization into our empirical analysis to account for potential confounding effects.

doing research, leading them to accept lower compensation for a stronger research orientation of their jobs (Sauermann and Roach 2013). It follows:

**Hypothesis 3:** Provided that the wage gap an academic scientist faces becomes smaller with more time spent on research (Hypothesis 1), this effect is larger for scientists with a high taste for science.

Finally, as mentioned above, the heterogeneity in scientists' tastes leads to strong self-selection in the scientific labor market, which has to be taken into account when wages of industrial and academic scientists are compared. Selection-effects in the labor market are comprehensively revealed in Roach and Sauermann (2010), who found that young Ph.D. students with a high taste for science are more likely to prefer a career in academe, while those with a greater concern for salary, access to resources, and applied research prefer a career in industry (see also Agarwal and Ohyama, 2013).

## 3. Methodological remarks

One of the fundamental problems in the empirical assessment of the academe-industry wage gap is the previously mentioned self-selection in the scientific labor market, which may distort a consistent estimation of the factual wage differential and its determinants. Since scientists are not randomly assigned to industry or academe but rather choose to do so based on their personal characteristics and preference for science, standard OLS regressions yield inconsistent estimates.

As non-random assignment is a very common problem in empirical policy evaluations, a number of methods that address this issue have been developed (see Blundell and Costa Dias for an overview of methods used in labor market policy evaluations, or Heckman et al., 1997, 1999;

Imbens and Wooldridge, 2009 for general overviews). We control for selection through a matching estimator.[5] Matching techniques have been used by several labor economists interested in wage gap estimations and labor policy evaluations (see, among others, Angrist, 1998; Card and Sullivan, 1988; Gerfin and Lechner, 2002; Frolich, 2007; Nopo, 2008). Matching has several advantages over parametric estimators (cf. Imbens and Wooldridge, 2009). It aims to create the conditions of an experiment with random assignment by constructing a control sample with equal characteristics as the treatment sample. While it does not impose a functional form on the selection, it relies on the assumption that selection takes place on observables, that is there are no unobserved variables that drive selection into academic or industrial positions. In order to minimize concerns in this regard we show below that our control sample consists only of industrial scientists that are indistinguishable from academic scientists along all known drivers of self-selection and several other personal characteristics.

The question that our empirical exercise attempts to answer is what an academic scientist with given characteristics would have earned in an industrial position, spending the same proportion of time on research and development oriented activities. Technically, we want to calculate the average treatment effect on the treated:

$$E(\alpha_{TT}) = E(Y_T|S = 1) - E(Y_C|S = 1),$$

where $S$ denotes whether the individual has been treated (1: in academia), $Y_T$ denotes the wage of the treated group (academic scientists), and $Y_C$ represents the wages of the control group

[5] Several other methods dealing with selection exists, including difference-in-difference designs, regression discontinuity designs, instrumental variable estimation, and control functions. The difference-in-difference method requires panel data on scientists that switch from academe to industry (or the other way around) due to an exogenous shock. As the data we employ is a cross-section and no exogenous shock is observed we cannot apply this estimator. A regression discontinuity design is also not applicable because there is no (arbitrary) rule that would determine the treatment (being an academic scientist). IV estimators as well as selection models rely on the idea to instrument selection into industry or academe through an instrumental variable, which affects the treatment but not the outcome. Valid instruments are, however, notoriously hard to find when it comes to wage regressions.

(industrial researchers). As we cannot directly observe the wages of academic researchers had they chosen to work in industry we have to estimate them. For this we rely on Rubin's Conditional Independence Assumption, which states that for a scientist with a set of exogenous characteristics $X$ participation and potential outcomes are independent from one another (Rubin, 1977):

$$E(Y_C|S = 1, X) = E(Y_C|S = 0, X)$$

As long as we are able to find for all academic scientists a counterpart in industry ($0 < P(S = 1|X) < 1$), which is known as the common support restriction, we can use the wages of those industrial scientists to estimate the wages of the academic scientists had they not been treated:

$$E(\alpha_{TT}) = E(Y_T|S = 1, X = x) - E(Y_C|S = 0, X = x).$$

In order to condition the samples on observable characteristics, we employ nearest neighbor matching (Heckman, Ichimura, and Todd, 1998), where we choose the most similar control observation for each treatment observation. However, matching on many dimensions makes it difficult to find appropriate controls. To overcome this issue we apply propensity score matching, where we estimate a propensity score $P(X)$ as single index, using a Probit model, and match on this score. Propensity score matching has no clear advantages or disadvantages over multidimensional matching (Rosenbaum and Rubin, 1983; Heckman, Ichimura, and Todd, 1998), but makes it easier to find adequate controls. The average treatment effect on the treated can finally be calculated as the mean difference of the matched samples:

$$\hat{\alpha}_{TT} = \frac{1}{n^T}\left(\sum_i Y_i^T - \sum_i \widehat{Y_i^C}\right),$$

with $\widehat{Y_i^C}$ being the counterfactual wage for scientist $i$ and $n^T$ being the sample size of the academic scientists.

## 4. Data

We employ data from the Belgian edition of the Careers of Doctorate Holders (CDH) survey (Federaal Wetenschapsbeleid, 2006). The survey took place in 2006 (reference year: 2005), and was created by the OECD in cooperation with Eurostat and the UNESCO Institute for Statistics. The primary goal was to gain detailed knowledge about career paths of doctorate holders.[6]

Compared to other countries in which the survey was carried out, the Belgian part contains more detailed questions, including selected ones about the scientists' research activities that allow a much more accurate approximation of activities than other large scale data sources (e.g. the frequently used National Science Foundation's Survey of Doctorate Recipients). Combined with information on scientists' and researchers' earnings, it offers an unique chance to investigate the academe-industry wage differential and how the nature of research activities affects it.

The response rate of the survey is estimated to be 17.7 percent (7,160 responses). Of these, we select natural scientists and engineers currently employed in industry or academe and being in the first 30 years of their careers. After removing observations with inconclusive answers (see appendix 1 for details), and retaining only those who answered all relevant questions, we arrive at a final sample covering 486 academic scientists and 759 industrial scientists. Below we describe all variables used in the analysis (see Table 1 for variable definitions).

[6] Detailed information on the project can be found in Unesco (2012), Eurostat (2012), or OECD (2012). Summary statistics based on the full survey have been published in Auriol (2007, 2010). Statistics on the Belgian data collection can be found in Moortgat & Van Mellaert (2011). For the methodological background and core questionnaire see Auriol et al. (2012).

## Table 1: Variable definitions

| Variable | Description |
|---|---|
| WAGE | Reported earnings before taxes in 2005, taking all occupations into account. |
| Motivations | |
| TASTE_SCIENCE | Respondents were asked to indicate why they chose to be a scientist (7 possible reasons in binary form). We applied exploratory factor analysis (results reported in Appendix 2), extracting two factors. TASTE_SCIENCE correlates positively with intellectual challenge, work circumstances, contribution to society, and independence, representing intrinsic motivation and higher preference for 'classic' science. |
| TASTE_COMMERCIALIZATION | TASTE_COMMERCIALIZATION correlates positively with salary, extralegal benefits, career prospects, and job security, representing extrinsic motivation. |
| Nature of research activities | |
| RES_SHARE | Respondents were asked how much of their time they spend on average on research and development activities. RES_SHARE is the sum of the share of time assigned to "executing or guiding research" and "interpreting others' research". |
| DEV_SHARE | DEV_SHARE is the sum of the share of time assigned to "improving existing products or processes" and "developing new products or processes". Both range from 0 to 100 percent of time. |
| General controls | |
| CAREER_YEARS | Time since graduation in years. |
| AGE | The age of the respondent in years. |
| TIME_CURRENTJOB | The time the respondent has been employed in the current job, in years. |
| RELATIONSHIP | 1 if the respondent indicates to be cohabitating or married, 0 otherwise. |
| CHILDREN | 1 if the respondent indicates to have children younger than 18, 0 otherwise. |
| GENDER | 1 if the respondent is female, 0 otherwise. |
| NATURAL SCIENCES | 1 if the respondent indicates to be active in the natural sciences ("mathematics, computer science, informatics"; physics; chemistry; geology; biology; other natural sciences), 0 if the respondents indicates to be active in engineering or technological sciences (civil, mechanical, chemical, electrical, electronic engineering; material sciences; medical engineering techniques; environmental science; biotechnology; nanotechnology; other engineering or technological disciplines). |
| HOURS_WORKED | Self-reported average number of hours worked per week. |
| Controls for ability | |
| TIMETOPHD_PROP | The time the respondent spent to complete his or her doctorate. Calculated as the deviation from the mean time to completion within detailed scientific field and ten-year graduation cohort. A score equal to one indicates that the doctorate was completed at the average time within field and cohort. A score lower than one indicates that the doctorate was completed more quickly. |
| SCHOLARSHIP | 1 if the respondent received a government or private grant to fund his or her doctorate, 0 otherwise. |

**Dependent variable: wage**

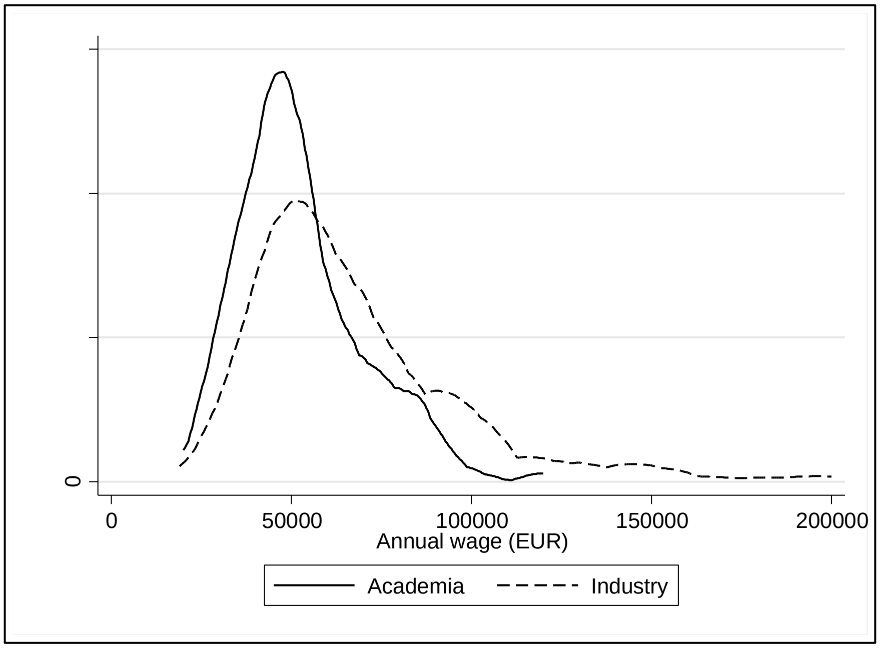
The main outcome of interest is the sum of the scientist's or researcher's self-reported annual earnings before taxes.[7,8] Academics reported earnings of 52,203 EUR on average, while industrial researchers reported 67,019 EUR, which amounts to a wage gap of approximately 29 percent. Median wages are 49,100 EUR for academics versus 60,000 EUR for industrial researchers. These numbers are very similar to the median wages reported by the Survey of Doctorate Recipients in the US, which were 65,000 USD for academics versus 85,000 USD for industrial researchers, or a wage gap of 30% (National Science Board, 2012). For a more precise assessment of the sectorial wage differences, Figure 2 plots the wage distributions within academe and industry. The earnings of industrial scientists are shifted somewhat to the right. They also show a significantly longer right tail than academic wages, reflecting flatter earnings profiles in academia.

Academic wage setting in Belgium differs structurally from the United States, as academics are usually hired as civil servants and the corresponding wage scales for public sector workers are set by a governmental decree in negotiation with strong labor unions. Wages in the Belgian academic sector are therefore only to small degree influenced by competition in the scientific labor market. The main source of variance in academic wages stems from job tenure, position and side activities, e.g. speaking fees, which are included in the CDH's definition of the yearly wage.

[7] The results do change only slightly when we take hourly wage as the main variable of interest instead of the annual wage. The corresponding results are presented in Appendix 3.

[8] In order to ensure a clear differentiation between academic and industrial scientists and control for systematic differences caused through multiple employments, we disregard scientists and researchers with multiple jobs.

**Figure 2: Wage distribution by sector**




### Explanatory variables

Activities: time spent on research and development

The CDH survey includes questions about how scientists spend their working time. These questions allow a more fine-grained measurement of scientists' activities compared to most statistics used in the extant literature, which typically indicate the main focus of a scientists' research activities, e.g. basic versus applied research, but do not provide a weighted measure of different activities (e.g. Agarwal and Ohyama, 2013; Sauermann and Stephan, 2012).

In order to approximate the proportion of time a scientist spends on research activities (RES_SHARE) we sum the share of time spent on "performing or guiding research" and "interpreting the research of others". The time a researcher spends on development oriented research (DEV_SHARE) is proxied by the total share of time spent on "improving existing products or processes" and "developing new products or processes".[9] Descriptive statistics show that academic scientists spend on average 50 percent of their time on research activities, and 7 percent of their time on development activities, while these numbers correspond to respectively 29 percent and 40 percent for industrial researchers. The picture that these numbers draw is in line with the common notion that academic scientists tend to focus more on basic rather than applied research, while industrial scientists tend to spend much more time on product and process development.

A crucial advantage of our data is the multidimensional continuous measurement of engagement in research and development. Many scientists spend nontrivial shares (≥1%) of their time on research as well as development: 30% of the academics, and 65% of researchers. Only 16% of academics indicate to be full-time researchers, and only 6% of industrial researchers are engaged full-time in development.

Preferences

The literature on scientists' motivations finds that scientists have a certain taste for science as well as commercialization (Roach and Sauermann, 2010; Sauermann and Roach, 2012; Agarwal and Ohyama, 2013). Taste for commercialization represents motivation through extrinsic factors, such as wages, extralegal benefits, and job security (for a definition of taste for science see section 2 and the cited literature). The CDH survey addresses scientists' motivations through the question "why did you choose for a career as a scientist?", followed by possible reasons, which were recorded in binary form. Table 2 lists the responses by sector. The

[9] Agarwal and Ohyama (2013) define scientists' activities as primarily basic or primarily applied in nature, while Sauermann and Stephan (2012) make a distinction between scientists primarily engaged in basic research, those primarily engaged in applied research, and those primarily engaged in development. Our measure of research activities is most likely mainly driven by basic research activities but also picks up applied research components.

differences are consistent with previous research on motivation-driven sector choice based on US scientists' data (Roach and Sauermann, 2010).

**Table 2: Motivations by sector**

| | Academia | Industry | p-value of two-sided t-test on mean differences |
|---|---|---|---|
| MOTIV_CHALLENGE | 0.84 | 0.62 | $p<0.0001$ |
| MOTIV_SALARY | 0.04 | 0.08 | $p=0.0106$ |
| MOTIV_BENEFITS | 0.00 | 0.02 | $p=0.0035$ |
| MOTIV_CAREER | 0.16 | 0.18 | $p=0.3046$ |
| MOTIV_SECURITY | 0.09 | 0.04 | $p=0.0039$ |
| MOTIV_WORKCIRC | 0.29 | 0.18 | $p<0.0001$ |
| MOTIV_INDEPENDENCE | 0.65 | 0.33 | $p<0.0001$ |
| MOTIV_CONTRIBUTION | 0.17 | 0.06 | $p<0.0001$ |

Notes: Share of respondents in each sector who indicated being motivated by factor.

In order to enhance comparability with the literature, tastes for science and commercialization of researchers are discerned based on exploratory factor analysis.[10] The analysis returned two independent factors representing the scientists' taste for science and commercialization (see Appendix 2 for detailed results). In line with descriptive statistics from the US, academic scientists score higher on taste for science than industrial scientists (0.43 versus -0.28, two-sided t-test on mean differences: $t(1243)=-13.07$, $p<0.001$), and lower on taste for commercialization (-0.12 versus 0.79, two-sided t-test on mean differences: $t(1243)=3.52$, $p<0.001$). Scientists and researchers can have a high taste for science or business at the same time (cf. Sauermann and Roach, 2012). It should be noted that tastes might change throughout time and are possibly subject to socialization in the workplace (Sauermann & Roach, 2012), while we measure them at the time the survey was carried out. As such, the results concerning preferences should be interpreted with care.

[10] Our results are robust to alternatively taking the sums of MOTIV_CHALLENGE, MOTIV_CONTRIBUTION, and MOTIV_INDEPENDENCE to represent taste for science.

How well matching works hinges on the conditioning factors that affect selection into treatment and outcome. We control for as many factors as possible in the analysis. In addition to preferences and research activities, we include scientists' ability, efforts, career experience and demographic characteristics as covariates. It might be argued that some of these factors do not directly cause selection, or are subject to reverse causality. For example, scientists potentially choose the amount of effort they put into their work as a function of wage prospects, and time spent on research and development are consequences of selection instead of causes. As such, the matching should not be viewed with a causal interpretation in mind. Nevertheless, it is crucial for the identification of the wage gap to condition on these factors, so we include all of them as covariates in the analysis.

Ability

Ability influences our estimations in two ways. First, more able scientists and researchers are likely to be more productive, and thus earn more. Second, ability has been shown to cause selection into applied or basic research: in academe, more able scientists are more likely to prefer basic research topics than applied research, since academe tends to offer more resources to conduct basic research. Higher complementarities between basic and applied research in industry cause these differences in access to resources to be smaller, leading to no such selection in industry (Agarwal and Ohyama, 2013). As ability is usually unobservable to the econometrician it is important to find suitable proxies that pick up as much as possible of the unobserved variance. Since all respondents have attained a Ph.D., the level of attained education, which is otherwise often used to proxy for ability, does not matter here. Indicators such as publication or citation counts are biased in the context of this study: institutional differences between industry and academe lead to academics being more prone to publish. Reliance on publications would thus lead to an upward ability bias in academics' ability as compared to industrial researchers. The CDH survey allows us to create two proxies that are not affected by selection into industry or academe. The first one is based on the main source of funding of the respondents' Ph.D. studies. Government and private Ph.D. scholarships tend to select strongly on applicants' previous academic performance.[11] At the same time, Ph.D. students are strongly encouraged to apply for these scholarships. Therefore, having a scholarship (SCHOLARSHIP) should separate a higher-skilled group from a lower-skilled group. Agarwal and Ohyama (2013) employ grant data in a similar fashion as one among two other controls for ability.

The time a Ph.D. holder spent to complete his dissertation serves as a second proxy for ability. In order to deal with heterogeneity across fields and time, we calculate time to completion as a proportional deviation from the mean time to completion within a field[12] and a ten-year graduation cohort (TIMETOPHD_PROP). A value equal to one indicates that the scientist completed the Ph.D. in an average amount of time, while values less than one indicate faster and thus better performance. A similar measure is used by Agarwal and Ohyama (2013), where time to completion of Bachelor's degree is taken.

Both proxies have the advantage of being independent of later career decisions, which is not the case with measures such as publication output. Instead, they are based on variation within the respondent's Ph.D. education. The main disadvantage is that there is some unobserved ability variance that we have to leave uncontrolled.

Effort

[11] For instance, one of the main Flemish institutions distributing Ph.D. scholarships (FWO) lists "research ability and potential (including course results)" and "research skills and methodology" as the first two selection criteria for its Ph.D. fellowships (FWO, 2013)
[12] We consider the researchers' detailed research domain (16 domains) for this calculation.

Effort is an important determinant of wage. Therefore, we include the self-reported number of hours worked during an average week as a further matching variable (HOURS_WORKED).[13]

Career Experience

Experience is another well-known driver of scientists' wages: several studies indicate that academic scientists' earnings profiles are concave from below, and peak late in the career. (Diamond, 1986; Laitner and Stafford, 1985; Lillard and Weiss, 1979; Weiss and Lillard, 1978; Creedy, 1988; Stephan, 1996). Age and cohort effects are also important for scientists' productivity (Stephan and Levin, 1992; Stephan, 1996), further affecting wages. In the matching, we control for three time-related factors. The first, AGE, is the natural age of the scientist. The second, CAREER_YEARS, is the time since a scientist obtained the Ph.D. The last time-related measure is the time spent in the current job (TIME_CURRENTJOB). As scientists spend more time in a position, they accumulate job-specific skills which probably increase the wage. In academe specifically, wages are known to rise with job tenure and academic rank, which both increase over time.

Demographics

Supplementary to the work related attributes we take some demographic factors into account that may have an influence on wage setting as well as career decisions. These include gender (GENDER, takes value one if the scientist is a woman), relationship status (RELATIONSHIP, takes value one if the scientist is married or cohabiting), and whether the scientist has children (CHILDREN, takes value one if the scientist has children younger than 18).

Field

[13] Appendix 3 provides results based on the hourly wage instead of the annual wage as an alternative to control for differences in scientists' efforts.

Finally, we include a dummy variable that indicates whether a scientist is mainly active in natural sciences, while the baseline category represents engineering. Table 3 lists summary statistics for all variables described above. Academic and industrial scientists differ systematically with regard to their tastes, research activities, job tenure, demographics and other characteristics. Table A3 in the appendix presents correlations.

## Table 3: Summary statistics

| Variable | Academia before matching (n=486) | | | | | Potential controls in industry (n=759) | | | | | p-value of two-sided t-test on mean differences |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | Std. Dev. | Min | Median | Max | Mean | Std. Dev. | Min | Median | Max | |
| TASTE_SCIENCE | 0.43 | 0.86 | -1.63 | 1.02 | 1.27 | -0.28 | 0.98 | -1.94 | -0.36 | 1.47 | p<0.001 |
| RES_SHARE | 50.34 | 35.67 | 0.00 | 50.00 | 100.00 | 29.18 | 28.98 | 0.00 | 20.00 | 100.00 | p<0.001 |
| DEV_SHARE | 7.43 | 15.63 | 0.00 | 0.00 | 100.00 | 40.16 | 32.33 | 0.00 | 40.00 | 100.00 | p<0.001 |
| TASTE_COMMERCIALIZATION | -0.12 | 0.43 | -0.55 | -0.29 | 2.05 | 0.08 | 1.23 | -0.55 | -0.14 | 9.60 | p<0.001 |
| CAREER_YEARS | 11.85 | 8.38 | 0.00 | 10.50 | 29.75 | 11.45 | 7.53 | 0.00 | 10.00 | 29.92 | p=0.381 |
| AGE | 41.70 | 8.75 | 27.00 | 41.00 | 63.00 | 40.88 | 7.48 | 27.00 | 39.00 | 61.00 | p=0.080 |
| TIME_CURRENTJOB | 8.84 | 9.23 | 0.00 | 5.00 | 37.00 | 7.41 | 6.77 | 0.00 | 5.00 | 30.00 | p=0.002 |
| RELATIONSHIP | 0.83 | 0.38 | 0.00 | 1.00 | 1.00 | 0.85 | 0.36 | 0.00 | 1.00 | 1.00 | p=0.436 |
| CHILDREN | 0.54 | 0.50 | 0.00 | 1.00 | 1.00 | 0.64 | 0.48 | 0.00 | 1.00 | 1.00 | p<0.001 |
| GENDER | 0.22 | 0.41 | 0.00 | 0.00 | 1.00 | 0.16 | 0.38 | 0.00 | 0.00 | 1.00 | p=0.011 |
| TIMETOPHD_PROP | 1.00 | 0.30 | 0.05 | 0.96 | 2.40 | 0.97 | 0.28 | 0.02 | 0.94 | 2.72 | p=0.176 |
| SCHOLARSHIP | 0.50 | 0.50 | 0.00 | 0.00 | 1.00 | 0.56 | 0.50 | 0.00 | 1.00 | 1.00 | p=0.023 |
| HOURS_WORKED | 50.10 | 8.38 | 38.00 | 50.00 | 72.00 | 48.11 | 7.03 | 38.00 | 48.00 | 72.00 | p<0.001 |
| NATURAL SCIENCES | 0.63 | 0.48 | 0.00 | 1.00 | 1.00 | 0.51 | 0.50 | 0.00 | 1.00 | 1.00 | p<0.001 |
| WAGE | 52203.12 | 17348.76 | 20000.00 | 49100.00 | 120000.00 | 67019.13 | 29556.26 | 19000.00 | 60000.00 | 200000.00 | p<0.001 |

Notes: Definitions of the variables are provided in Table 1.

# 5. Results

## 5.1. Basic results of the matching

Table A4 presents the Probit model used in the matching. For 29 of the academic scientists, no adequate match could be found due to lack of common support. Hence, we exclude them from the following investigation. As this reduction amounts to less than 3 percent of the full sample and the sample mean wage and other controls stay almost the same (no statistically significant changes), the sample reduction does not affect the analysis. No significant differences remain between the two groups after the matching, as shown by two-sided t-tests on difference in the means of each variable in Table 4. Further, Table A4, column 2, provides the results of a re-estimation of the aforementioned Probit model based on the matched sample of academic and industrial scientists. A $\chi^2$-test on the joint significance of all explanatory variables indicates that both samples show no significant differences in their covariates anymore ($\chi^2(14)=6.10$, $p=0.9640$).

# Table 4: Summary statistics after matching

| Variable | Academia after matching (n=456) | | | | | Selected controls in industry (n=456) | | | | | p-value of two-sided t-test on mean differences[#] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | Std. Dev. | Min | Median | Max | Mean | Std. Dev. | Min | Median | Max | |
| TASTE_SCIENCE | 0.40 | 0.87 | -1.63 | 1.02 | 1.27 | 0.42 | 0.84 | -1.34 | 1.02 | 1.47 | p=0.830 |
| RES_SHARE | 47.78 | 34.95 | 0.00 | 50.00 | 100.00 | 50.23 | 41.80 | 0.00 | 52.50 | 100.0 | p=0.619 |
| DEV_SHARE | 7.89 | 16.02 | 0.00 | 0.00 | 100.00 | 9.37 | 14.95 | 0.00 | 0.00 | 90.00 | p=0.429 |
| TASTE_COMMERCIALIZATION | -0.11 | 0.44 | -0.55 | -0.29 | 2.05 | -0.13 | 0.47 | -0.55 | -0.29 | 7.06 | p=0.783 |
| CAREER_YEARS | 12.04 | 8.32 | 0.00 | 10.75 | 29.75 | 12.24 | 8.56 | 0.00 | 10.63 | 29.67 | p=0.877 |
| AGE | 41.79 | 8.70 | 27.00 | 41.00 | 63.00 | 42.16 | 8.56 | 0.00 | 1.00 | 1.00 | p=0.755 |
| TIME_CURRENTJOB | 8.69 | 9.16 | 0.00 | 5.00 | 37.00 | 8.53 | 7.50 | 0.00 | 6.50 | 30.00 | p=0.849 |
| RELATIONSHIP | 0.83 | 0.37 | 0.00 | 1.00 | 1.00 | 0.82 | 0.38 | 0.00 | 1.00 | 1.00 | p=0.823 |
| CHILDREN | 0.57 | 0.50 | 0.00 | 0.00 | 1.00 | 0.58 | 0.49 | 0.00 | 1.00 | 1.00 | p=0.836 |
| GENDER | 0.22 | 0.41 | 0.00 | 0.00 | 1.00 | 0.16 | 0.36 | 0.00 | 0.00 | 1.00 | p=0.199 |
| TIMETOPHD_PROP | 0.99 | 0.30 | 0.05 | 0.96 | 2.40 | 1.00 | 0.29 | 0.43 | 0.95 | 2.72 | p=0.912 |
| SCHOLARSHIP | 0.52 | 0.50 | 0.00 | 1.00 | 1.00 | 0.46 | 0.50 | 0.00 | 0.00 | 1.00 | p=0.393 |
| HOURS_WORKED | 49.82 | 8.22 | 38.00 | 50.00 | 70.00 | 50.51 | 8.02 | 38.00 | 50.00 | 70.00 | p=0.502 |
| NATURAL SCIENCES | 0.62 | 0.49 | 0.00 | 1.00 | 1.00 | 0.63 | 0.48 | 0.00 | 1.00 | 1.00 | p=0.860 |
| WAGE | 52474.41 | 17382.71 | 20000.00 | 49800.00 | 120000.00 | 64302.80 | 33311.52 | 21339.00 | 54000.00 | 200000.00 | p=0.004 |

Notes: #: p-values calculated using Lechner's (2001) approximation of standard errors to account for sampling with replacement. 29 observations are excluded from the analysis after matching because the sample is restricted to common support as described in section 3. Definitions of the variables are provided in Table 1.

After matching, the average treatment effect on the treated (the difference between the average annual wage in industry and the average annual wage in academe) becomes €11.828 (two-sided t-test on mean differences of wages before and after matching: $t(455)=-2.91$, $p=0.004$), or a decrease of the wage gap by 20.2 percent (initially €14.816). In absolute numbers, the wage gap decreased by 5.8 percentage points from 28.3 percent to 22.5 percent.[14] This result suggests that 20 percent of the wage gap is due to differences between industry and academe in the time spent on research and development activities, tastes for science, and other demographic characteristics. Academic scientists thus face on average a lower wage gap than descriptive statistics suggest.

Even though differences in research activities and preferences are believed to be the main drivers of the academe-industry wage differential, the wage gap remains significant when these structural differences are controlled for. This could have several reasons. First, there might be significant differences in the appropriation of the returns to research between industry and academia. Firms are generally more inclined to commercialize findings, while academic institutes are more inclined to publish them. The differing goals of research might lead a researcher in industry to generate more appropriable value than the same researcher in academia. While the empirical analysis controls for research activities, this remains a structural difference between academic and industrial research. Second, academic scientists and researchers in industry might pursue different activities outside of research. Academics mostly teach or participate in the administration of the university when not doing research or development, whereas industrial researchers might engage in management activities, which generate more value for the firm. While we implicitly control for the share of time the researcher spends on non-research activities (through the shares of time spent on research and development), this

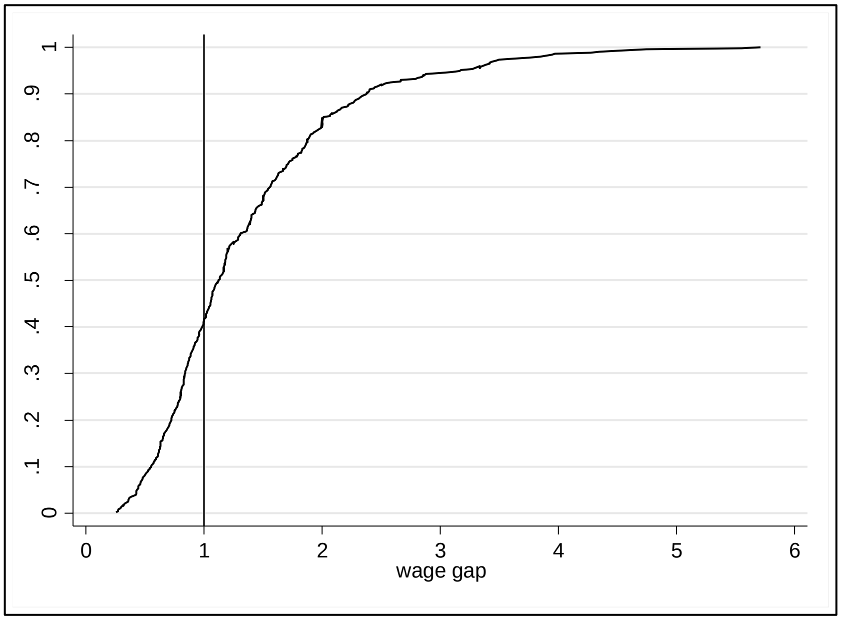
[14] Note that all results are even more pronounced if one considers the median wages instead of the means.

remains another structural difference. Third, the expectation that the wage gap closes after matching builds on the belief that industrial scientists are willing to forgo monetary rewards for a higher focus on research activities, while academic scientists' wages are relatively unaffected by the time they spend on research. The relatively small decrease of the wage gap after matching might indicate that other wage determinants that differ systematically across academic and industrial researchers exert a positive influence on the wage differential, such as differences in the time spent on development orientated research activities. In order to disentangle the positive and negative factors underlying the wage gap, the upcoming section presents regressions of the estimated wage differences on research activities and other covariates.

Figure 4 illustrates the heterogeneity in the wage gap through a plot of the cumulative distribution of the relative difference between the observed wage in academe and the estimated counterfactual wage in industry. While the average wage gap after matching is 22.5%, 40% of the sampled academics actually earn more than the estimated wage in industry.[15] On the other hand, almost 20% of the sample might earn at least twice as much money if they would work in industry.

[15] It should be noted here that this evaluation of the wage gap assumes that the academic would work for the average firm, instead of matching to a firm which suits his needs. If this matching were to happen, the wage gap could become even smaller for scientists who spend much time on research, as they are more likely to sort into firms which offer high freedom and lower wages. Scientists who spend more time on development should be more likely to sort into companies which focus more strongly on development, leading to an even higher wage differential.

**Figure 3: Cumulative distribution of the wage gap**




## 5.2. Explaining the wage gap

The aim of following analysis is to examine how differences in the time spent on research and development influence the academe-industry wage gap, holding personal characteristics constant. Therefore, we run a standard OLS regression of the previously estimated wage gap per academic scientist on the scientist's time spent on research, time spent on development, taste for science, taste for commercialization, years since graduation in linear and quadratic terms, time in the current job, demographic controls, controls for ability, average hours worked, and a field control.[16] In order to reduce the skewness of the wage gap distribution we estimate a semi-elastic model, taking the natural logarithm of the relative wage differential as the dependent variable. Table 5 presents the results.

[16] To avoid multicollinearity we include CAREER_YEARS but not AGE in the regression. The results are robust to employing AGE instead of CAREER_YEARS.

## Table 5: Regression results

| | (1) | | (2) | | (3) | | (4) | | (5) | |
|---|---|---|---|---|---|---|---|---|---|---|
| Dependent variable | | | $\ln\left(\frac{Wage\ in\ industry}{Wage\ in\ academia}\right)$ | | | | $\ln(wage)$ in Academia | | $\ln(wage)$ of selected industry controls | |
| | Coef | Se | Coef | Se | Coef | Se | Coef | Se | Coef | Se |
| TASTE_SCIENCE | -0.0162 | (0.0273) | -0.0508 | (0.0767) | -0.5060 | (0.3520) | 0.0011 | (0.0156) | -0.0396 | (0.0360) |
| RES_SHARE | -0.0032*** | (0.0008) | -0.0018 | (0.0011) | -0.0046*** | (0.0011) | 0.0006 | (0.0004) | -0.0023*** | (0.0008) |
| DEV_SHARE | 0.0030** | (0.0013) | 0.0034* | (0.0019) | 0.0043** | (0.0020) | 0.0001 | (0.0006) | 0.0028* | (0.0015) |
| TASTE_COMMERCIALIZATION | 0.0052 | (0.0487) | -0.1715* | (0.0988) | -0.0302 | (0.0981) | 0.0369 | (0.0299) | 0.0016 | (0.0529) |
| CAREER_YEARS | -0.0315** | (0.0125) | 0.0083 | (0.0191) | -0.0498*** | (0.0159) | 0.0309*** | (0.0065) | 0.0856*** | (0.0163) |
| CAREER_YEARS_2 | 0.0002 | (0.0004) | -0.0009 | (0.0006) | 0.0009 | (0.0006) | -0.0003 | (0.0002) | -0.0024*** | (0.0006) |
| TIME_CURRENTJOB | -0.0042 | (0.0039) | -0.0010 | (0.0057) | -0.0098* | (0.0059) | 0.0003 | (0.0019) | 0.0218*** | (0.0065) |
| RELATIONSHIP | -0.0694 | (0.0761) | -0.0722 | (0.1196) | -0.0521 | (0.0930) | -0.0026 | (0.0415) | 0.3474*** | (0.0913) |
| CHILDREN | 0.0518 | (0.0575) | 0.0532 | (0.0879) | 0.0576 | (0.0761) | 0.0314 | (0.0268) | -0.0400 | (0.0707) |
| GENDER | 0.0688 | (0.0690) | 0.2842*** | (0.1031) | -0.0713 | (0.0885) | -0.0761** | (0.0323) | 0.0227 | (0.0768) |
| SCHOLARSHIP | 0.1034** | (0.0515) | 0.0551 | (0.0740) | 0.1234* | (0.0718) | -0.0287 | (0.0245) | -0.0452 | (0.0560) |
| TIMETOPHD_PROP | -0.0679 | (0.0754) | -0.0962 | (0.1224) | -0.0747 | (0.1018) | -0.0040 | (0.0376) | -0.1353 | (0.1041) |
| HOURS_WORKED | -0.0055* | (0.0031) | 0.0003 | (0.0041) | -0.0101** | (0.0042) | 0.0053*** | (0.0015) | -0.0002 | (0.0047) |
| NATURAL | 0.0260 | (0.0513) | 0.0890 | (0.0742) | -0.0042 | (0.0688) | -0.0621*** | (0.0235) | -0.0758 | (0.0634) |
| INTERCEPT | 0.9240*** | (0.2195) | 0.2172 | (0.3044) | 1.8725*** | (0.4451) | 10.2601*** | (0.1159) | 10.3016*** | (0.2592) |
| Number of observations | 456 | | 174 | | 282 | | 456 | | 456 | |
| $R^2$ | 0.226 | | 0.254 | | 0.280 | | 0.473 | | 0.628 | |
| Root MSE | 0.5170 | | 0.4645 | | 0.5357 | | 0.2389 | | 0.2977 | |
| F-statistic | 10.6699 | | 4.8140 | | 8.5379 | | 29.6733 | | 11.4601 | |

Notes: This table presents OLS regressions. Heteroscedasticity robust standard errors in parentheses. Stars indicate significance levels of coefficients: *: p<0.10; **: p <0.05; ***: p <0.01. (1) dependent variable: treatment effect (ln(wage in industry/wage in academe)), full sample. (2) treatment effect regression for scientists with taste for science lower than or equal to the mean. (3) treatment effect regression for scientists with taste for science larger than the mean. (4) dependent variable: ln(wage) of academic scientists. (5) dependent variable: ln(wage) of selected industrial researchers. Clustered standard errors take repeated observations caused by selection without replacement into account. Definitions of the variables are provided in Table 1.

The evidence supports the hypothesis that the wage gap is smaller for scientists who spend more time on research: it declines by 0.32 percent for each additional percent of time spent on research (column 1). Consistent with previous findings (e.g. Stern, 2004) and the notion of a dominating preference effect (hypothesis 1), differences in research orientation between academic and industrial scientists explain a large part of the positive wage differential. Comparing only academic scientists with those industrial scientists that are able to spend similar amounts of time on research, leaving all other covariates unaccounted, the academe-industry wage gap would indeed be much smaller than descriptive statistics suggest. Academic scientists who are full-time engaged in research would earn only 1.6% more in industry, while the average wage gap is 22.5%. In line with our conceptualization of wage dynamics in industry and academe, this effect is driven by a negative relation between research and wage in industry, while academic wages are not sensitive to differences in research time (models 4 and 5, Table 5).

Spending more time on development increases the wage gap. With each additional percent of time spent on development oriented research by academic scientists, a comparable industrial scientist's wage increases by 0.3 percent. The magnitude of the effect is not trivial: an average academic scientist who spends 50 percent of his time on development oriented research is subject to an estimated wage differential of 73.5 percent.[17,18] Wage regressions within both sectors (models 4 and 5, Table 5) further show that this effect stems largely from a positive relation between development oriented research and wages within industry, while the relation is insignificant within academe. In terms of preference versus productivity effects, the negative relation between research and wages can be interpreted as preference effects dominating productivity effects, in line with the findings of Stern (2004) and Sauermann and Roach (2013).

[17] 23 of the academics in the sample, or 5%, spend at least 50% of their time on development.
[18] As noted above, these predictions assume that scientists go to the average firm, instead of sorting into firms according to their preferences.

That development relates positively to industrial wages could either indicate a higher productivity effect: there occurs more rent-sharing with scientists who focus on more profitable development, or a lower preference effect: scientists enjoy working on development less and demand higher wages to do so. While our data do not allow us to separate these alternatives, we do show that the relation between wage and science is not exclusively negative: research in its most applied form relates to higher wages in industry.

Apart from differences in research activities the second most important factor affecting the wage gap is years after graduation: the wage gap decreases by approximately 3.2 percent per year.[19] This result is in line with earlier findings of Stevens (2004) and Agarwal and Ohyama (2013), reporting converging earnings paths of academic and industrial scientists in the US. This effect seems to be driven by diminishing wage increases throughout time in industry, compared to linear wage increments in academe (column 4 and 5, Table 5).

To test the moderating role of taste for science on the relation between research orientation and the wage gap (hypothesis 3), we split the sample in a low-taste-for-science and high-taste-for-science group based on the mean values (columns 2 and 3).[20] In line with the conceptual reasoning, the impact of RES_SHARE on the wage differential is higher in the high-taste-for-science group (column 3) than in the low-taste-for-science group (column 2), indicating that high taste for science researchers are willing to trade of wage for more research orientation at higher levels than low taste for science researchers (F-test on differences between the coefficients of RES_SHARE in model 2 and 3: $F(1,426)=3.12$, $p=0.0780$). The results imply that academic

[19] Since the turning point of the estimated relationship between career years and the relative wage gap lies far out of the observed range of career years (315 years), the influence is almost linear with only slightly decreasing marginal effects. Including only the linear term or taking a logarithmic specification does not alter the results.

[20] Alternatively, one could include a corresponding interaction term in the regression. However, interaction term specifications imply the assumption that all other covariates exert the same effect within both sub-groups, which is in our case not true.

researchers with a relatively high taste for science who spend much time on research face much lower comparable wages in industry. All other covariates equal, the wage gap for scientists with a rather high taste for science becomes zero at 86 percent of time spent on research. Scientists with a high taste for science who want to spend most of their time on research might thus be financially best off in an academic position. The effect of time spent on development is also larger for scientists with a high taste for science compared to those with a rather low taste for science, but the difference is not statistically significant (0.0034 in the low group versus 0.0043 in the high group, $F(1,426)=0.11$, $p=0.7460$).

Although all findings are in line with empirical investigations based on US scientists' data, it is important to keep in mind that the Belgian academic sector is strongly regulated. Consequently, our findings might be limited in their generalizability to more competitive academic environments. Academic researchers working in less regulated markets might be better able to commercialize their research findings, leading to a positive relation of research activities and their wages in academe. This could in turn influence the wage gap. There might also be a positive relation between scientific output and wages in academe, leading to an even smaller wage gap for highly able academic scientists who focus on basic research.

A final limitation of our analysis is that the applied proxies of research and development activities are rather broad and the actual underlying activities might still be quite heterogeneous. Although being more detailed than those measures used in previous studies, even more detailed information is needed to further disentangle which kinds of research activities, such as the possibility to publish research findings, are mainly responsible for wage differences between both sectors. 'Research' and 'Development' might also carry different meanings in academe and industry. While the survey clearly stated what is meant by each activity, divergent interpretations between the sectors remain a concern.

## 6. Conclusion

This paper provides new evidence on the academe-industry wage gap. The analysis reveals that the wage differential varies largely with individual scientists' focus on research tasks as opposed to development activities. Academic scientists who spend much time on research face a much lower counterfactual wage in a comparable position in industry, whereas academic scientists who spend much time on development activities face a much higher one. How much scientists pay to be scientists hinges apparently much on their research orientation.

We further find that a scientist's taste for science does not seem to influence the wage gap directly, but scientists with a relatively high taste for science trade off monetary rewards against more research focus at higher levels than scientists with a relative low taste for science. Our results are in line with the notion that in terms of 'research' activities preference effects dominate productivity effects, whereas in terms of 'development' activities productivity effects dominate preference effects.

While differences in research activities and preferences explain a large part of the wage gap, significant differences remain. Probable reasons for the remaining wage gap are a better appropriation and monetization of research findings in industry and differences in non-research related activities among academics and researchers, e.g. teaching versus management.

Our results have practical implications for academic institutions, policy makers, and individual scientists. For academic institutions it is worthwhile to note that offering scientists the freedom to pursue their own research projects compensates for much lower monetary rewards in academe. Urging scientists to focus on more applied research projects might cause an unwanted selection of academics into industry since scientists can expect to receive much higher returns to development oriented research in industrial positions. Alleviating the commercialization of academic scientist's research findings might a be way to address this potential source of an academic brain drain. If commercialization is preferred, academic institutes might also put more weight on successful research commercialization when it comes promotion decisions. That way, engaging in applied research can have positive wage benefits in academe, which would narrow the wage gap.

From a policy point of view the relatively low wage gap of research focused scientists and the rather large wage gap of development oriented scientists might ensure an efficient allocation of different types of scientists into both sectors. Politicians and managers of academic institutions should, however, be aware of the fact that the less time academic scientists are able to focus on research the more attractive it becomes for scientists to seek an industrial position.

Our findings also have a bearing on individual researchers' decision whether to pursue a career in academe or industry. As getting into the academic labor market becomes harder, scientists could consider employment in research-like industrial jobs. While those jobs offer lower wages compared to development-focused industrial research jobs, the wage difference compared to academe turns out to be rather small. Hence, they might be considered as welcome outlets for scientists who are forced to leave academe but still want to pursue research. It is also interesting to know that the perceived higher wages in industry are conditional on a focus on development oriented research and other activities that are unrelated to research. Future investigations might be able to answer the question which role the individual wage gap plays for a scientist's career decision and how it influences the allocation of research talent into both sectors.

## References

Agarwal, R., Ohyama, A. (2013). Industry or academia, basic or applied? Career choices and earnings trajectories of scientists. Management Science, 59, 950-970.

Aghion, P., Dewatripont, M., Stein, J. (2008). Academic freedom, private-sector focus, and the process of innovation. Rand Journal of Economics, 39(3), 617-635.

Altbach, P., Reisberg, L., Yudkevich, M., Androushcak, G., Pacheco, I. (2012). Paying the professoriate: a global comparison of compensation and contracts. London: Routledge.

Angrist, J. (1998). Estimating the labor market impact of voluntary military service using social security data on military applicants. Econometrica, 66(2), 249-288.

Arora, A., Gambardella, A. (1994). The changing technology and technological change: general and abstract knowledge and the division of innovative labour. Research Policy, 23(5), 523-532.

Auriol, L. (2007). Labour market characteristics and international mobility of doctorate holders: results for seven countries. OECD Science, Technology and Industry Working Papers 2007/02. Retrieved from http://dx.doi.org/10.1787/310254328811.

Auriol, L. (2010). Careers of doctorate holders: employment and mobility patterns. OECD Science, Technology and Industry Working Papers 2010/04. Retrieved from http://dx.doi.org/10.1787/5kmh8phxvvf5-en.

Auriol, L., Schaaper, M., Felix, B. (2012). Mapping careers and mobility of doctorate holders: draft guidelines, model questionnaire and indicators. OECD Science, Technology and Industry Working papers 2007/6. Retrieved from http://dx.doi.org/10.1787/5k4dnq2h4n5c-en.

Barbezat, D. (1987). Salary differentials by sex in the academic labour market. Journal of Human Resources, 22, 422-428.

Bayer, A., Astin, H. (1968). Sex differences in academic rank and salary among science doctorates in teaching. Journal of Human Resources, 2, 191-200.

Blundell, R., Costa Dias, M. (2009). Alternative approaches to evaluation in empirical microeconometrics. Journal of Human Resources, 44, 565-640.

Card, D., Sullivan, D. (1988). Measuring the effect of subsidized training-programs on movements in and out of employment. Econometrica, 56, 497-530.

Cohen, W., Levinthal, D. (1990). Absorptive capacity: a new perspective on learning and innovation. Administrative Science Quarterly, 35, 128-152.

Creedy, J. (1988). Cohort and cross-sectional earnings profiles: scientists in Britain and Australia. Journal of Economic Studies, 15(1), 44-52.

Dasgupta, P., David, P. (1994). Towards a new economics of science. Research Policy, 23, 487-521.

Diamond, A. (1986). The life-cycle research productivity of mathematicians and scientists. Journal of Gerontology, 41, 520-525.

Eurostat (2012). Careers of doctorate holders. Retrieved 05 03, 2012, from http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Careers_of_doctorate_holders

Federaal Wetenschapsbeleid. (2006). Careers of Doctorate Holders Survey [Database]. Gent: ECOOM UGent.

Fini, R., Lacetera, N. (2010). Different yokes for different folks: individual preference, institutional logics, and the commercialization of academic research. In G. Libecap (Ed.),

Spanning boundaries and disciplines: university technology commercialization in the idea age (Vol. 21, pp. 1-25). Bingley: Emerald Group Publishing.
Frolich, M. (2007). Propensity score matching without conditional independence assumption - with an application to the gender wage gap in the United Kingdom. Econometric Journal, 10, 359-407.
FWO. (2013). Ph.D. Fellowship. Retrieved 01 27, 2013, from FWO: www.fwo.be/Aspirant.aspx.
Gerfin, M., Lechner, M. (2002). A microeconometric evaluation of the active labour market policy in Switzerland. Economic Journal, 112, 854-893.
Heckman, J., Ichimura, H., Todd, P. (1998). Matching as an econometric evaluation estimator. Review of Economic Studies, 65, 261-294.
Heckman, J., Ichumura, H., Todd, P. (1997). Matching as an econometric evaluation estimator: evidence from evaluating a job training program. Review of Economic Studies, 64, 605-654.
Imbens, G., Wooldridge, J. (2009). Recent developments in the econometrics of program evaluation. Journal of Economic Literature, 47, 5-86.
Karlsson, M., Trygg, L., Elfström, B.-O. (2004). Measuring R&D productivity: complementing the picture by focusing on research activities. Technovation, 24, 179-186.
Konrad, A., Pfeffer, J. (1990). Do you get what you deserve? Factors affecting the relationship between productivity and pay. Administrative Science Quarterly, 35, 258-285.
Lacetera, N. (2009). Different missions and commitment power in RD organizations: theory and evidence on industry-university alliances. Organization Science, 20, 565-582.
Laitner, J., Stafford, F. (1985). The academic labor market: has compensation diverged from other professions? Econometric Society Meetings. Washington, DC.
Lillard, L., Weiss, Y. (1979). Components of variation in panel earnings data: American scientists. Econometrica, 47, 437-454.
McNabb, R., Wass, V. (1997). Male-female salary differentials in British universities. Oxford Economic Papers, 49, 328-343.
Merton, R. (1973). The sociology of science: theoretical and empirical investigations. Chicago: University of Chicago Press.
Moore, W., Newman, R., Turnball, G. (1998). Do academic wages decrease with seniority? Journal of Labour Economics, 16, 352-366.
Moortgat, P., Van Mellaert, G. (2011). CDH (Careers of Doctorate Holders). Onderzoek, ontwikkeling en innovatie in België studiereeks 12 (Available at http://www.belspo.be/belspo/organisation/Publ/pub_ostc/ind/ind12_nl.pdf ed.). Brussels, Belgium: Belgian Science Policy Office.
National Science Board. (2012). Science and Engineering Indicators 2012. Arlington VA: National Science Foundation.
Nopo, H. (2008). Matching as a tool to decompose wage gaps. Review of Economics and Statistics, 90, 290-299.
OECD (2013). OECD/UNESCO Institute for Statistics/Eurostat Careers of Doctorate Holders (CDH) project. Retrieved 08 06, 2013, from http://www.oecd.org/innovation/inno/oecdunescoinstituteforstatisticseurostatcareersofdoctorateholderscdhproject.htm
Ong, L., Mitchell, J. (2000). Professors and hamburgers: an international comparison of relative academic salaries. Applied Economics, 32, 869-876.

Roach, M., Sauermann, H. (2010). A taste for science? PhD scientists' academic orientation and self-selection into research careers in industry. Research Policy, 39, 422-434.
Rosenbaum, P., Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70, 41-55.
Rosenberg, N. (1990). Why do firms do basic research (with their own money)? Research Policy, 19, 165-174.
Rubin, D. (1977). Assignment to a treatment group on the basis of a covariate. Journal of Educational Statistics, 2, 1-26.
Sauermann, H., Roach, M. (2013). Not all scientists pay to be scientists: heterogeneous preferences for publishing in industrial research. Research Policy, forthcoming.
Sauermann, H., Roach, M. (2012). Taste for science, taste for commercialization, and hybrid scientists. Paper presented at DRUID conference, June 19-21, Copenhagen Business School, Denmark.
Sauermann, H., Stephan, P. (2012). Conflicting logics? A multidimensional view of industrial and academic science. Organization Science, Articles in Advance, 1-21.
Stephan. (1996). The economics of science. Journal of Economic Literature, 34, 1199-1235.
Stephan, P., Levin, S. (1992). Striking the mother lode in science: the importance of age, place, and time. Oxford: Oxford University Press.
Steunpunt tot bestrijding van armoede, bestaansonzekerheid en sociale uitsluiting (2012). Feiten en cijfers. Retrieved 15 02, 2012, from http://www.armoedebestrijding.be/cijfers_minimum_uitkeringen.htm
Stern, S. (2004). Do scientists pay to be scientists? Management Science, 50, 835-853.
Stevens, P. (2004). Academic salaries in the UK and US. National Institute Economic Review, 190, 104-113.
Uebersax, J. (2000). Estimating a latent trait model by factor analysis of tetrachoric correlations. Retrieved 03 16, 2011, from http://john-uebersax.com/stat/irt.htm.
Unesco (2012). Tracking the careers of doctorate holders. Retrieved 07 05, 2012, from http://www.uis.unesco.org/ScienceTechnology/Pages/doctorate-degree-holders.aspx
Walker, J., Vignoles, A., Collins, M. (2010). Higher education academic salaries in the UK. Oxford Economic Papers, 62, 13-35.
Weiss, Y., Lillard, L. (1978). Experience, vintage, and time effects in the growth of earnings: American scientists. Journal of Political Economy, 86, 427-474.

## Appendix 1: Data

Starting from the initial dataset of the Careers of Doctorate Holders (CDH) survey we discarded a number of observation for various reasons. First, 53 academic scientists were removed from the sample, because they reported distinctly non-academic job descriptions, including ‘webmaster’, ‘secretary’, ‘manager’, or ‘coordinator’. Second, 38 academic scientists were removed who reported being employed in an academic and a non-academic position at the same time. Those who reported wages below the Belgian annual minimum wage (€17.689) and those above €200.000 (the 99$^{th}$ percentile) were not included in the analysis. The latter group is so rare and unique in the Belgian academic labor market that it would be meaningless to search for comparable scientists and would give them an unwarranted strong influence on the average wages. In order to further ensure a reasonable comparability of the scientists included in the analysis, the sample was further restricted to scientists and researchers in the first 30 years after graduation, who were older than 18 and younger than 65 at the moment the survey was carried out. We also removed scientists and researchers who reported working less than full-time (38 hours) or more than 75 hours a week (the 99$^{th}$ percentile). Scientists and researchers who reported spending in total more than 100 percent of their time on research and development were re-scaled to reflect a maximum time allocation of 100 percent.

## Appendix 2: Exploratory Factor Analysis used in tastes calculation

Since the items are all binary, we employ the Tetrachoric correlation matrix for this analysis (Uebersax, 2000).[21] The analysis returns a two factor solution, explaining 87 percent of total variation. Several rotations were applied from which we select the Varimax solution for interpretability. The first factor correlates with motivation through salary, extralegal benefits, career prospects, job security, and weakly with work circumstances. We name this factor "Taste for Commercialization". The second factor correlates with motivation through intellectual challenge, independence, contribution to society, and, more strongly, with work circumstances. We name this factor "Taste for Science". Both were normalized for interpretability.

**Table A1: Tetrachoric correlation matrix**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1. MOTIV_CHALLENGE | 1.00 | | | | | | | |
| 2. MOTIV_SALARY | 0.24 | 1.00 | | | | | | |
| 3. MOTIV_BENEFITS | 0.19 | 0.73 | 1.00 | | | | | |
| 4. MOTIV_CAREER | 0.42 | 0.47 | 0.36 | 1.00 | | | | |
| 5. MOTIV_SECURITY | 0.25 | 0.58 | 0.70 | 0.45 | 1.00 | | | |
| 6. MOTIV_WORKCIRC | 0.42 | 0.41 | 0.42 | 0.23 | 0.33 | 1.00 | | |
| 7. MOTIV_INDEPENDENCE | 0.79 | 0.13 | -0.04 | 0.18 | 0.28 | 0.51 | 1.00 | |
| 8. MOTIV_CONTRIBUTION | 0.59 | 0.18 | 0.34 | 0.13 | 0.36 | 0.21 | 0.45 | 1.00 |

**Table A2: Results of EFA**

| | Factor loadings | | Uniqueness |
|---|---|---|---|
| Item | Taste for commercialization | Taste for science | |
| MOTIV_CHALLENGE | 0.17 | **0.89** | 0.18 |
| MOTIV_SALARY | **0.78** | 0.12 | 0.38 |
| MOTIV_BENEFITS | **0.92** | 0.02 | 0.16 |
| MOTIV_CAREER | **0.51** | 0.24 | 0.68 |
| MOTIV_SECURITY | **0.77** | 0.22 | 0.35 |
| MOTIV_WORKCIRC | 0.40 | **0.47** | 0.62 |
| MOTIV_INDEPENDENCE | 0.00 | **0.92** | 0.15 |
| MOTIV_CONTRIBUTION | 0.25 | **0.54** | 0.65 |
| Eigenvalue | 2.55 | 2.28 | |
| % of variance | 46% | 41% | |

Note: Varimax rotation. Factor loadings over .40 appear in bold.

[21] Similar methods have been employed by Sauermann & Roach (2012).

## Table A3: Correlation matrix

| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | TASTE_SCIENCE | 1.00 | | | | | | | | | | | | | | | |
| 2 | RES_SHARE | 0.24* | 1.00 | | | | | | | | | | | | | | |
| 3 | DEV_SHARE | -0.15* | -0.23* | 1.00 | | | | | | | | | | | | | |
| 4 | TASTE_COMMERCIALIZATION | 0.03 | 0.01 | 0.02 | 1.00 | | | | | | | | | | | | |
| 5 | CAREER_YEARS | 0.01 | -0.12* | -0.12* | 0.07* | 1.00 | | | | | | | | | | | |
| 6 | AGE | 0.00 | -0.13* | -0.11* | 0.05 | 0.95* | 1.00 | | | | | | | | | | |
| 7 | TIME_CURRENTJOB | -0.03 | -0.07* | -0.07* | 0.06* | 0.73* | 0.73* | 1.00 | | | | | | | | | |
| 8 | RELATIONSHIP | 0.00 | 0.04 | -0.04 | 0.03 | 0.08* | 0.07* | 0.05 | 1.00 | | | | | | | | |
| 9 | CHILDREN | 0.02 | -0.01 | 0.03 | 0.02 | -0.03 | -0.02 | -0.06* | 0.38* | 1.00 | | | | | | | |
| 10 | GENDER | -0.03 | 0.07* | -0.13* | -0.05 | 0.13* | -0.14* | -0.08* | -0.11* | -0.15* | 1.00 | | | | | | |
| 11 | TIMETOPHD_PROP | -0.01 | -0.04 | -0.01 | -0.03 | -0.02 | 0.09* | 0.06* | 0.01 | 0.00 | 0.03 | 1.00 | | | | | |
| 12 | SCHOLARSHIP | 0.03 | 0.07* | 0.01 | 0.00 | 0.04 | -0.06* | -0.05 | 0.03 | 0.01 | 0.02 | -0.16* | 1.00 | | | | |
| 13 | HOURS_WORKED | 0.04 | 0.01 | -0.06* | 0.02 | 0.2* | 0.21* | 0.16 | 0.10* | 0.12* | -0.20* | -0.01 | -0.02 | 1.00 | | | |
| 14 | NATURAL SCIENCES | 0.00 | 0.07* | -0.10* | 0.04 | 0.11* | 0.06* | 0.03 | -0.04 | -0.03 | 0.07* | -0.02 | 0.07* | -0.09* | 1.00 | | |
| 15 | WAGE | -0.11 | -0.14* | 0.07* | 0.11* | 0.59* | 0.56* | 0.46* | 0.13* | 0.07* | -0.16* | -0.04 | 0.02 | 0.27* | -0.01 | 1.00 | |
| 16 | HIGHER | 0.33* | 0.28* | -0.50* | -0.09* | 0.04 | 0.05 | 0.08* | -0.02 | -0.07* | 0.07* | 0.03 | -0.04 | 0.11* | 0.10* | -0.27* | 1.00 |

Notes: *: significant at p<0.05. Variable definitions are provided in Table 1.

## Table A4: Propensity score regressions

| | (1) | | (2) | |
|---|---|---|---|---|
| Dependent variable: Treatment group | Before matching | | After matching | |
| TASTE_SCIENCE | 0.4295*** | (0.0478) | 0.0031 | (0.0753) |
| RES_SHARE | 0.0071*** | (0.0013) | -0.0012 | (0.0022) |
| DEV_SHARE | -0.0314*** | (0.0021) | -0.0033 | (0.0035) |
| TASTE_COMMERCIALIZATION | -0.2121*** | (0.0763) | 0.0562 | (0.1202) |
| CAREER_YEARS | -0.0754*** | (0.0204) | -0.0009 | (0.0310) |
| AGE | 0.0495** | (0.0197) | -0.0078 | (0.0277) |
| TIME_CURRENTJOB | 0.0294*** | (0.0085) | 0.0088 | (0.0123) |
| RELATIONSHIP | -0.0579 | (0.1325) | 0.1077 | (0.2062) |
| CHILDREN | -0.2045** | (0.0982) | -0.0204 | (0.1594) |
| GENDER | 0.0657 | (0.1159) | 0.2338 | (0.1720) |
| SCHOLARSHIP | -0.1285 | (0.0945) | 0.1338 | (0.1508) |
| TIMETOPHD_PROP | 0.0110 | (0.1676) | 0.0426 | (0.2512) |
| HOURS_WORKED | 0.0227*** | (0.0060) | -0.0040 | (0.0107) |
| NATURAL | 0.2867*** | (0.0930) | -0.0406 | (0.1491) |
| INTERCEPT | -2.4178*** | (0.6555) | 0.3478 | (1.0027) |
| N | 1245 | | 914 | |
| Pseudo-$R^2$ | 0.38 | | 0.01 | |
| Test of joint significance | $\chi^2(14)$= 630.52*** | | $\chi^2(14)$=6.10 | |

Notes: This table reports Probit model estimations of being in the treatment group. Standard errors in parentheses. After matching: clustered standard errors to account for repeated observations. * p<0.10, ** p <0.05, *** p <0.01. Variable definitions are provided in Table 1.

## Appendix 3: Hourly wage results

| | (1) | | (2) | | (3) | | (4) | | (5) | |
|---|---|---|---|---|---|---|---|---|---|---|
| Dependent variable | | | $\ln\left(\frac{Wage\ in\ industry}{Wage\ in\ academia}\right)$ | | | | $\ln(wage)$ in academia | | $\ln(wage)$ of selected industry controls | |
| | Coef | Se | Coef | Se | Coef | Se | Coef | Se | Coef | Se |
| TASTE_SCIENCE | -0.0455* | (0.0269) | -0.0580 | (0.0743) | -0.5824* | (0.3501) | -0.0003 | (0.0156) | -0.0409 | (0.0356) |
| RES_SHARE | -0.0040*** | (0.0008) | -0.0022* | (0.0012) | -0.0059*** | (0.0011) | 0.0006 | (0.0004) | -0.0023*** | (0.0008) |
| DEV_SHARE | 0.0040*** | (0.0013) | 0.0035* | (0.0018) | 0.0060*** | (0.0019) | 0.0001 | (0.0006) | 0.0027* | (0.0015) |
| TASTE_COMMERCIALIZATION | 0.0334 | (0.0499) | -0.1576* | (0.0894) | 0.0024 | (0.0990) | 0.0381 | (0.0297) | 0.0011 | (0.0521) |
| CAREER_YEARS | -0.0303** | (0.0125) | 0.0069 | (0.0185) | -0.0491*** | (0.0159) | 0.0304*** | (0.0066) | 0.0856*** | (0.0161) |
| CAREER_YEARS_2 | 0.0003 | (0.0004) | -0.0009* | (0.0006) | 0.0010* | (0.0006) | -0.0002 | (0.0002) | -0.0023*** | (0.0006) |
| TIME_CURRENTJOB | -0.0067* | (0.0039) | -0.0003 | (0.0052) | -0.0149** | (0.0061) | 0.0002 | (0.0019) | 0.0214*** | (0.0064) |
| RELATIONSHIP | -0.0452 | (0.0769) | -0.0681 | (0.1176) | -0.0087 | (0.0924) | -0.0026 | (0.0417) | 0.3402*** | (0.0910) |
| CHILDREN | 0.0554 | (0.0580) | 0.0536 | (0.0818) | 0.0775 | (0.0775) | 0.0322 | (0.0267) | -0.0378 | (0.0701) |
| GENDER | 0.0604 | (0.0699) | 0.2620** | (0.1007) | -0.0715 | (0.0903) | -0.0740** | (0.0322) | 0.0184 | (0.0757) |
| SCHOLARSHIP | 0.1396*** | (0.0509) | 0.0619 | (0.0705) | 0.1783** | (0.0706) | -0.0289 | (0.0245) | -0.0431 | (0.0555) |
| TIMETOPHD_PROP | -0.1008 | (0.0764) | -0.1618 | (0.1137) | -0.0986 | (0.1053) | -0.0021 | (0.0373) | -0.1198 | (0.1025) |
| HOURS_WORKED | 0.0129*** | (0.0031) | 0.0198*** | (0.0040) | 0.0069* | (0.0041) | -0.0142*** | (0.0015) | -0.0195*** | (0.0047) |
| NATURAL | -0.0025 | (0.0514) | 0.0761 | (0.0708) | -0.0457 | (0.0691) | -0.0605** | (0.0234) | -0.0697 | (0.0630) |
| INTERCEPT | 0.0505 | (0.2205) | -0.6337** | (0.2934) | 1.0918** | (0.4536) | 3.4662*** | (0.1157) | 3.4830*** | (0.2571) |
| Number of observations | 456 | | 174 | | 282 | | 456 | | 456 | |
| $R^2$ | 0.238 | | 0.304 | | 0.288 | | 0.434 | | 0.647 | |
| Root MSE | 0.5158 | | 0.4446 | | 0.5380 | | 0.2388 | | 0.2947 | |
| F-statistic | 10.6697 | | 5.8267 | | 8.0369 | | 26.9804 | | 10.0561 | |

Notes: This table presents OLS regression results. Heteroskedasticity robust standard errors in parentheses. Stars indicate significance levels of coefficients: *: p<0.10; **: p <0.05; ***: p <0.01. wage = wage per hour. (1) dependent variable: treatment effect (ln(wage in industry/wage in academia)), full sample. (2) treatment effect regression for scientists with taste for science lower than or equal to mean. (3) treatment effect regression for scientists with taste for science higher than mean. (4) dependent variable: ln(wage) of academic scientists. (5) dependent variable: ln(wage) of selected industrial researchers. Clustered standard errors to take repeated observations caused by selection without replacement into account. Definitions of the variables are provided in Table 1.