# Overconfidence and Gender Differences in Wage Expectations

Stephanie Briel* Gregor Pfeifer*†‡§ Mirjam Reutter* Sascha Satlukal*¶

February 1, 2019

**Abstract**

We analyze the impact of overconfidence on gender differences in wage expectations using elicited beliefs of German university applicants. Interestingly, female students have lower wage expectations and are less overconfident than their male counterparts. Oaxaca-Blinder decompositions show that a substantial part (7.6%) of the gender gap in wage expectations can be explained by stronger overconfidence of males. Applying recentered influence function decompositions, we find that the impact of overconfidence on the gender gap is particularly strong at the bottom and top of the wage expectation distribution.

JEL Codes: J16; D84; D91; C21

Keywords: Gender Pay Gap, Wage Expectations, Overconfidence

**Preliminary Draft**
please do not cite or circulate

*University of Hohenheim
†University College London
‡CESifo
§IZA
¶Corresponding author: University of Hohenheim, Department of Economics, Stuttgart, Germany; sascha.satlukal@uni-hohenheim.de

# 1 Introduction

The difference in pay between men and women is one of the most intensively investigated phenomenon in Economics. The fact that men on average earn more than women has been found in numerous studies over the last decades, as the meta-analysis of Weichselbaumer and Winter-Ebmer (2005) shows. Although the gender wage gap has been declining over the last decades in economically advanced countries, there still exists a considerable gender difference in earnings (Blau and Kahn, 2017). Most studies show that a substantial part of the gender wage gap is caused by differences between men and women in terms of occupational choice and human capital measures. More recently, differences in personality traits are also increasingly considered as potential explanations for income differences across gender (e.g. Nyhus and Pons, 2012; Blau and Kahn, 2017). Nonetheless, it usually remains a non-negligible residual of the gender gap that cannot be explained by observed characteristics and is often interpreted as discrimination against women. The largest part of the gender wage gap, however, is most often attributed to the fact that men and women sort differently into occupations and industries.

Why do men work in different occupations and industries than women? For high-skilled individuals, the choice of occupation and industry is closely linked to their college major choice. Similarly to gender differences in occupational choice, men and women also sort differently into college majors (Zafar, 2013). Males, for instance, are more likely to study in so-called STEM fields (Science, Technology, Engineering, and Mathematics) than females are (Osikominu and Pfeifer, 2018). This suggests that a part of the gender wage gap for high skilled individuals is already caused by college major choice, since there exist large earning differences across college majors (Arcidiacono, 2004). Human capital theory suggests that individuals choose level and type of education based on expected market returns to education (Betts, 1996). Indeed, several studies back this claim and find that earning expectations are an important determinant of college major choice (e.g. Arcidiacono et al., 2017; Wiswall and Zafar, 2015). Wiswall and Zafar (2015), for example, show that expected earnings do significantly affect college major choice of students, using an experimentally generated panel of beliefs. These findings show that students indeed consider their expected own future earnings when choosing their college major.

Moreover, expected own future earnings of students are also a good predictor for their realized earnings some years later. Webbink and Hartog (2004) confront expected starting salaries of Dutch students with their realized starting salaries four years later. They find no systematic difference between expected starting salaries and later realizations. However, their key finding is that differences in the expected starting salary by type of study and by various other individual characteristics are quite similar to the respective differences in realizations. This indicates that students are able to precisely predict their own starting salary and that determinants of their expectations hardly differ from determinants of their actual starting salaries. There might be several reasons for this. First of all, most students probably have at least a vague idea of average starting salaries in occupations related to their field of study. Hence, it is not surprising that the field of study similarly affects expected starting salaries of students and later realizations. Furthermore, students might realistically evaluate their abilities relative to other students and consider their preferences for non-pecuniary aspects of jobs when forming their earning expectations. Students with high abilities might not only expect higher starting salaries but also earn more at labor market entry. Contrarily, students that have strong preferences for non-pecuniary aspects of jobs, such as flexible working hours, might have lower expectations and choose a job that is less well paid than students with weak preferences for non-pecuniary aspects. Additionally, it can also be the case that expected starting salaries of students become self-fulfilling. Whereas students with high income expectations are more likely to negotiate for a higher salary so that their expectations are satisfied, students with low income expectations are more likely to accept a less-paying job since it matches their beliefs (Reuben et al., 2017). There are also indications that expectations of students affect their academic performance, which in turn affects their realized earnings (e.g. Jacob and Wilder, 2011).

Analyzing gender differences in wage expectations can be very helpful to understand why men earn on average more than women, since realized wages are affected by wage expectations trough several channels and the structures of expectations and realizations are similar. Reuben et al. (2017) even argue that looking at expected earnings may be more informative for the explanation of gender differences in education and career choices than looking at realized earnings, since the latter might be affected by unanticipated future events and might suffer from the problem of reversed causality. In contrast to the gender gap in realized earnings, there are relatively few studies on gender differences in earning expectations. This might be in part due to the rather historical reluctance of economists to work with subjective expectations data, stemming from the doubt whether elicited subjective expectations of survey participants reflect their true expectations—given that there is nothing at stake for them. Botelho and Pinto (2004) show that these concerns are unwarranted. They conduct an experiment that is designed to elicit students' beliefs about economic returns to college education. Students taking part in the experiment are given financial incentives for accurately reporting. They compare these elicited beliefs to beliefs elicited using hypothetical surveys without financial incentives and find no significant difference. Thus, they conclude that students do tend to respond meaningfully to questions regarding their earnings expectations. This result might be one reason why the use of subjective expectations data in Economics has increased over the last years (e.g. Arcidiacono et al., 2017; Baker et al., 2018; Reuben et al., 2017; Zafar, 2013).

In this paper, we use subjective expectations data from a survey at Saarland University, Germany, to investigate the gender difference in wage expectations. During the application process in 2011 and 2012, prospective students were asked to report expectations of their own future salary. Alongside wage expectations, students state their field of study and provide information about their personal and family background. These information can be used to study the determinants of students' wage expectations. The main focus of our study lies on the role of overconfidence in explaining wage differences. Prospective students did also state their expectation regarding the salary of average other students in the same field of study. Overconfidence of students can thus be measured by comparing the expected own salary with the expected salary of average others given the same college major. We examine how much of the gender gap in wage expectations is caused by gender differences in overconfidence.

The previous literature on confidence shows that people are often overconfident in their own abilities (Croson and Gneezy, 2009). Moreover, many studies find differences between males and females with regard to overconfidence. Barber and Odean (2001), for instance, analyze trading behavior of male and female investors. They find that men trade considerably more than women and thereby reduce their returns. From this they conclude that male investors are more overconfident than female ones. A similar result in a different context is from Bengtsson et al. (2005). They compare confidence of male and female students by making use of a particular design of an economics exam at Stockholm University. In order to get the best possible grade in the exam, students had to answer an additional question. However, answering the additional question correctly was only meaningful for the grade if a student performed well throughout all other questions. Although male students did not perform better than female ones in other questions, male students were more inclined to answer the additional question. The study of Soll and Klayman (2004) suggests that both male and female judges are overconfident, but also indicates that men are more overconfident than women. Niederle and Vesterlund (2007) examine whether men and women of the same ability differ in their selection into a competitive environment on the basis of a laboratory experiment. They find that men are more likely to select a competitive environment, since they are substantially more overconfident about their relative performance than women are. However, similarly to the results of Soll and Klayman (2004), their findings indicate that women tend to be overconfident as well. Dahlbom et al. (2011), contrarily, find that girls are even underconfident regarding their mathematics performance. They asked 14-year-old high school students in Sweden prior to a mathematics test what grade they thought they would get and compared it to actual grades after. Whereas female students were too pessimistic, male students were too optimistic. Gender differences in overconfidence, however, is not present in all settings. Nekby et al. (2008) analyze overconfidence of men and women in a running competition in Sweden, which is a competitive male-dominated setting. Runners taking part in the competition chose their starting position based on the running time they expect to need. This choice is then compared with the running time from the previous year and the realized time in the same race. Their results suggest that female runners are even slightly more overconfident than male ones. Similarly, Hardies et al. (2013) study overconfidence in a highly selected and highly socialized group of professionals, namely external auditors, and compare it to the overconfidence of Belgian students. They find that male students are more overconfident than female ones, however, they find no gender difference regarding overconfidence in the group of auditors. To summarize, the previous literature suggests that men are more overconfident than women on average in most settings. However, as soon as women self-select into competitive, male-dominated environments, this difference seems to disappear.

By investigating the contribution of overconfidence to the gender gap in wage expectations, we combine three research areas: gender differences in pay, wage expectations, and overconfidence. Closest to our study is the work of Reuben et al. (2017), who analyze the impact of overconfidence and preferences for competitiveness and risk on earnings expectations of students. They use an experiment to measure such three factors and combine corresponding figures with data from a survey eliciting students' beliefs about their future income. Their estimations show that overconfidence and competitiveness do have a positive impact on the expected own future salary. However, the preference for risk does not. Furthermore, they also study gender differences in expected earnings and how this difference is influenced by their experimental measures. They find that men have considerably higher earnings expectations compared to women, and that this difference is increasing in age.[1] A substantial part of this gender gap can be attributed to their experimental measures of overconfidence and competitiveness.

In contrast to Reuben et al. (2017), however, we go beyond simple linear regressions and analyze potential explanations of the gender difference in wage expectations and the role of overconfidence in more detail by applying Oaxaca-Blinder decompositions. One advantage of the Oaxaca-Blinder approach is that it allows to determine, in a way that is path independent, how much of the gender gap in wage expectations can be explained by gender differences in a certain characteristic (Firpo et al., 2011). Moreover, we also explicitly examine the gender gap and the contribution of overconfidence at different parts of the wage expectation distribution. It is often found that the gender wage gap is larger at the bottom (sticky floor effect) and at the top (glass ceiling effect) of the wage

[1] Reuben et al. (2017) ask students what they expect to earn at age 30 and at age 45.

distribution (Christofides et al., 2013). These effects might already occur in case of wage expectations. In order to elaborate on this, we use the method proposed by Firpo et al. (2018), which is based on recentered influence functions (RIFs). The gender gap at any quantile of the distribution can be decomposed in analogy the standard Oaxaca-Blinder approach, after respective recentered influence functions are computed. In addition, we conduct our analysis for different subgroups of students to examine whether there are heterogeneous effects. Another advantage of our study, compared to Reuben et al. (2017) and other studies investigating wage expectations, is that we can make use a relatively large sample.

In line with previous research, we find that male students expect higher future salaries than female students do. Moreover, we also find that overconfidence has a significantly positive effect on wage expectations. As was found previously, our measurement of overconfidence suggests that males are more overconfident than females. These results already indicate that overconfidence partly explains the different earnings expectations of males and females. Indeed, results of the Oaxaca-Blinder decomposition show that a substantial part of the gender gap in wage expectations is caused by the fact that male students are more overconfident. Comparing the gender gap in wage expectations at different quantiles of the distribution, we do not find large variations. However, we do find that the contribution of overconfidence to the gender gap is larger at the bottom and at the top of the wage expectation distribution than in middle.

The remainder of the paper is organized as follows. Section 2 describes the data, shows first descriptive statistics, and explains our overconfidence measure. In Section 3, we describe the methods applied to decompose the gender gap in wage expectations. Results of the estimations are shown in Section 4. We first report simple linear regression estimates to get an impression of the relationship between overconfidence and the wage expectations, as well as of the size of the gender gap. Subsequently, we show results of the Oaxaca-Blinder decomposition with a particular focus on the overconfidence measure. Next, we focus on decomposing the gender gap at every decile of the wage expectations distribution and we outline decomposition results for several subgroups of students. Section 5 reports results of robustness checks, and Section 6 concludes.

# 2 Data and Descriptives

## 2.1 Data Set

The data we use is based on a student survey at Saarland University, Germany. Students applying for enrollment in academic years 2011 and 2012 were widely surveyed about their beliefs regarding starting salaries conditional on field of study. The data were first used by Klößner and Pfeifer (2018), who evaluate characteristics causing German students to make larger or smaller estimation errors in the prediction of future salaries after adjusting wage expectations for misconceptions of the German income tax.[2] Whereas in 2011 only 500 students completed the survey, the number increased to 1,561 students in 2012 since two further fields of study were added (Education and Medicine). So, overall the sample consists of 2,061 students, which is a relatively large data set compared to other studies examining students wage expectations (e.g. Reuben et al., 2017; Webbink and Hartog, 2004).[3]

Moreover, as Klößner and Pfeifer (2018) show, Saarland University and its students constitute a good representation of the average university and the average student body in Germany. Figures like student/teacher ratio, male/female ratio, student age distribution, distribution of graduates across fields, distribution of gender across fields, number of exams passed, grades, duration of studies, etc., are all close to the German average. Saarland University does also appear in the middle of international rankings across items such as Teaching/Learning, Research, Knowledge Transfer, International Orientation, or Regional Engagement.

The prospective students first had to state in which field of study and for which degree (Bachelor, Master, or State Examination) they had currently applied for. They were also asked whether they are planning to do a further degree afterwards (Master, Second State Examination, or Doctoral Degree) and with which degree they would aim to earn their first salary. In the second part of the survey, the prospective students were asked to answer several questions regarding monthly gross salaries.[4] First, they had to state their expectations of their own salary at labor market entry and after five years on the job, referring to the degree with which they intend to earn their first salary. Second, they were asked to estimate the respective salaries for average others within the same field of study and with the same degree.

Furthermore, the students were asked about several characteristics of their personal and family background. They should provide the following information: their gender; their age; their working experience; their final grade in secondary school; whether their mother and father graduated from college and, if so, in which major discipline; whether they intended to live at their parents' house while studying; whether they expect to receive "BAfoeG"[5] and, if so, how much; their school system in secondary school; and the federal

[2]Throughout this study, we focus on wage expectations that are adjusted for students' misconceptions of the progressive income tax.

[3] We drop 13 prospective students since they have missing values in their wage expectations.

[4] Students were provided with detailed explanations about the difference between gross and net salaries.

[5] The Bundesausbildungsförderungsgesetz (BAfoeG) is a Federal Training Assistance Act that regulates federate student grants and loans in Germany. It supports students from a weaker financial

state in which they obtained their higher education entrance qualification.

Finally, students should also answer two questions regarding the importance of their future income: they were asked how influential the income expectation has been on their college major choice and how important it is to receive an above-average salary.

## 2.2 Descriptive Statistics

Table 1 depicts the means and standard deviations of wage expectations and several other relevant variables for male and female students, respectively. The last column of Table 1 also shows t-statistics regarding a test of equality of means between males and females.

Male students have on average significantly higher expectations w.r.t. their starting salary as well as w.r.t. the salary they earn after five years on the job than female students. Males expect on average a starting gross salary of 3,579 € and a gross salary after five years of 5,540 €. Females, in comparison, expect to earn on average just 3,015 € at labor market entry and 4,652 € after five years. Both male and female students plausibly expect their earnings to rise with increasing working experience. Moreover, male students do not only have significantly higher expectations with regard to their own future earnings but also with regard to the earnings of average others within the same field of study. Hence, males do expect higher wages for themselves and a higher overall wage level in their field of study than females do. There is also a gender difference in the relation between the expected own salaries and the expected salaries of average others. Whereas the means of own expected salaries are higher than those of expected average salaries for male students, the opposite is the case for female students. Table 11 in the appendix shows the adjusted wage expectations differentiated by field of study. Mathematics and Computer Science students expect the highest starting salaries for themselves and for average others. Law Students, however, expect the highest earnings after five years on the job. Education and Humanities students, contrarily, feature the lowest income expectations.

Male students in the sample are on average 21.22 years of age, which is slightly older than female students, who are on average 20.78 years of age. There is no significant gender difference with respect to working experience. However, male students did perform worse in secondary school on average than their female counterparts. The final grade in secondary school of males is on average 2.31, whereas that of females is only 2.16.[6]

Moreover, males and females differ substantially with regard to the choice of the field of study, as is also found by the previous literature (e.g. Zafar, 2013). Males in the sample are significantly more likely to apply for a program in Business Studies as well as in Mathematics and Computer Science than females are. 16 percent of males, for instance,

background.

[6] In Germany, grades are scaled from 1 to 6, with 1 being the best grade, and 6 being the worst.

Table 1: Descriptive Statistics

| | Male | Female | $t$-value |
|---|---|---|---|
| Wage Expectations | | | |
| Own Starting Salary | 3579.12 (1908.16) | 3014.89 (1558.45) | 7.14 |
| Average Starting Salary | 3571.00 (1839.61) | 3178.37 (1620.37) | 5.02 |
| Own Salary After Five Years | 5539.82 (3262.18) | 4652.17 (2654.29) | 6.58 |
| Average Salary After Five Years | 5339.48 (3276.44) | 4767.36 (2688.77) | 4.21 |
| Human Capital | | | |
| Age | 21.22 (3.85) | 20.78 (3.50) | 2.61 |
| Final Grade Secondary School | 2.31 (0.60) | 2.16 (0.59) | 4.91 |
| Working Experience | 4.36 (17.30) | 4.23 (16.22) | 0.17 |
| Field of Study Shares (Percent) | | | |
| Business Studies | 29.77 (45.75) | 19.61 (39.72) | 5.25 |
| Education | 6.32 (24.35) | 11.21 (31.56) | -3.95 |
| Humanities | 9.43 (29.23) | 15.03 (35.75) | -3.90 |
| Law Studies | 12.18 (32.73) | 14.09 (34.81) | -1.27 |
| Math./Comp. Science | 16.21 (36.87) | 3.57 (18.55) | 9.28 |
| Medicine | 18.85 (39.13) | 29.03 (45.41) | -5.43 |
| Natural Sciences | 7.24 (25.93) | 7.47 (26.30) | -0.20 |
| Influence of Income Expectation on Choice of Field of Study (Percent) | | | |
| Very Low | 26.67 (44.25) | 28.44 (45.13) | -0.89 |
| Low | 17.13 (37.70) | 18.34 (38.71) | -0.71 |
| Neutral | 37.93 (48.55) | 37.01 (48.30) | 0.42 |
| Strong | 16.09 (0.37) | 14.43 (35.16) | 1.03 |
| Very Strong | 2.18 (14.62) | 1.78 (13.24) | 0.64 |
| Importance of an Above-Average Salary (Percent) | | | |
| Very Unimportant | 4.94 (21.69) | 5.52 (22.84) | -0.58 |
| Unimportant | 8.51 (27.91) | 11.88 (32.37) | -2.53 |
| Neutral | 35.29 (47.81) | 42.44 (49.45) | -3.00 |
| Important | 42.30 (49.43) | 35.14 (47.76) | 3.28 |
| Very Important | 8.97 (28.59) | 5.01 (21.82) | 3.41 |
| Observations | 870 | 1178 | |

*Note:* First tow columns show means and standard deviations (in parenthesis). Last column shows $t$-values of a test of equality of means between males and females. Wage expectations are expected gross salaries that were adjusted for misconceptions of the progressive income tax by Klößner and Pfeifer (2018).

are applying for a program in Mathematics and Computer Science compared to only 4 percent of females. Contrarily, a significantly larger share of females is applying for a program in Education, Humanities, and Medicine. For example, whereas 29 percent of females do apply for a program in Medicine, only 19 percent of males do.

There are also differences between male and female students in terms of the view on the importance of income. An above-average salary seems to be more important for male students. 9 percent of males, for instance, state that an above-average salary is very important compared to only 5 percent of females. 42 percent of males compared to 35 percent of females regard an above-average income as important. This indicates that males are more strongly driven by pecuniary incentives than females are and that females have stronger preferences for non-pecuniary aspects of jobs, which is also found, amongst others, by Wiswall and Zafar (2018). Males and females differ less, though, with regard to the question how influential the potential income was on their college major choice.

## 2.3 Overconfidence Measure

Our data allows to compare the expectations of students for their own starting salary, conditional on field of study, with their expectations for the average starting salary of a student in the same field of study. This can be used to construct a measurement of the individual overconfidence of a respective student. For the analysis the following overconfidence index is used:

$$oc_{it} = \frac{y_{own,it}}{y_{other,it}} \qquad \text{with} \quad t = \text{start, after five years.} \tag{1}$$

Which is simply the ratio of the expected own salary of individual $i$, $y_{own,it}$, to the average salary of all other fellow students, expected by individual $i$, $y_{other,it}$. Thus, the index takes on a value larger than one if the student expects to earn more than average, which we consider as overconfidence. Similarly, it takes on a value smaller than one if the student expects to earn less than others, which could be interpreted as underconfidence. Consequently, the index is equal to one if a student expects to earn as much as her fellow students.

Means of the overconfidence index for males and females are shown in Table 2. Males are on average more overconfident than females, both w.r.t. the starting salary as well the salary after five years. The differences between males and females are statistically significant as the test of equality of means shows in the last row. This is in line with most of the previous literature on overconfidence (e.g. Barber and Odean, 2001; Bengtsson et al., 2005; Niederle and Vesterlund, 2007; Dahlbom et al., 2011). The mean of males is larger than one for both salaries, , i.e., male students on average expect to earn more

Table 2: Mean of the Overconfidence Index by Gender

| | $oc_{start}$ | $oc_{five}$ | $t$-value |
|---|---|---|---|
| Male | 1.061 | 1.101 | -1.294 |
| | (0.624) | (0.671) | |
| Female | 0.983 | 1.016 | -1.348 |
| | (0.325) | (0.773) | |
| Overall | 1.016 | 1.052 | -1.865 |
| | (0.477) | (0.732) | |
| $t$-value | 3.353 | 2.658 | |

*Note:* Means and standard deviations (in parenthesis) of the respective overconfidence index. The last column shows $t$-values of a test of equality of means between the overconfidence index with regard to the starting salary and with regard to the salary after five years. The last row shows $t$-values of a test of equality of means between males and females.

than the expected average salary at the start of their working career (6 percent) as well as five years after (10 percent). Contrarily, females expect to earn on average 2 percent less than the expected average starting salary. However, they are more confident with respect to the salary they earn after five years. Here, they expect to earn on average 2 percent more than average others. The difference in overconfidence with regard to the starting salary and the salary after five years is insignificant for males and females separately, but it is weakly significant in the full sample as the test statistics in the last column show. A reason for this difference could be that overconfident students expect their presumed superior ability or higher working effort to be more influential on their own salary after they have proven their skills to their employer for a few years.

Even though the mean of the overconfidence index is larger than one for males (both salaries) and females (salary after five years), it is not the case that the majority of students are overconfident. In fact, a larger share of both male and female students expects to earn even less than the expected average starting salary, which can be seen in Table 3. Nevertheless, 23 percent of males and 16 percent of females are overconfident with respect to the starting salary. The largest share of males (47 percent) and females (46 percent) expects the same starting salary as the starting salary of their fellow students. Regarding the salary after five years, the shares of males and females that are overconfident are larger compared to the starting salary. There are more males that are overconfident than males who are underconfident. However, the share of overconfident females is still smaller than the share being underconfident. Once again, the largest share of male and female students expects to earn the average salary, i.e., showing an overconfidence ratio of one. However, the shares are somewhat smaller for the salary after five years compared to the starting salary.

Table 4 reports differences in the mean overconfidence indices with respect to the field of study a student applied for. In most fields, males are on average more overconfident

Table 3: Overconfidence Shares

| | $oc_{start}$ | | | $oc_{five}$ | | |
|---|---|---|---|---|---|---|
| | $> 1$ | $= 1$ | $< 1$ | $> 1$ | $= 1$ | $< 1$ |
| Male | 23.45 | 46.67 | 29.89 | 32.30 | 40.69 | 27.01 |
| Female | 16.21 | 45.50 | 38.29 | 20.12 | 44.06 | 35.82 |
| Overall | 19.29 | 46.00 | 34.72 | 25.29 | 42.63 | 32.08 |

*Note:* Table shows the respective shares in percent.

Table 4: Mean of the Overconfidence Index by Gender and Field of Study

| | $oc_{start}$ | | $oc_{five}$ | |
|---|---|---|---|---|
| Field of Study | Male | Female | Male | Female |
| Business Studies | 1.035 | 0.980 | 1.057 | 0.988 |
| | (0.389) | (0.217) | (0.357) | (0.192) |
| Education | 0.980 | 1.009 | 0.995 | 1.013 |
| | (0.105) | (0.418) | (0.133) | (0.331) |
| Humanities | 0.975 | 0.985 | 1.040 | 0.963 |
| | (0.243) | (0.378) | (0.306) | (0.231) |
| Law Studies | 1.007 | 0.959 | 1.058 | 0.982 |
| | (0.336) | (0.392) | (0.345) | (0.272) |
| Math./Comp. Science | 1.007 | 0.958 | 1.033 | 0.979 |
| | (0.239) | (0.198) | (0.190) | (0.165) |
| Medicine | 1.268 | 0.993 | 1.341 | 1.019 |
| | (1.267) | (0.321) | (1.395) | (0.366) |
| Natural Sciences | 1.023 | 0.967 | 1.052 | 1.272 |
| | (0.253) | (0.187) | (0.216) | (2.636) |
| Overall | 1.061 | 0.983 | 1.101 | 1.016 |
| | (0.624) | (0.325) | (0.671) | (0.773) |

*Note:* Table shows means and the standard deviations (in parenthesis) of the respective overconfidence index.

than females. However, a few exceptions exist. Male Education students, for instance, are less confident than their female counterparts. They expect to earn on average slightly less than the average salary, while female Education students expect to earn slightly more than the average salary. Females who study Natural Sciences are also more overconfident than males with regard to the salary after five years. Comparing all fields of study shows that male Medicine students are by far the most overconfident subgroup with regard to both salaries. On average, they expect to earn 27 percent more than the average salary at the start of their working career, and 34 percent more than average five years later. Female students of Natural Sciences are very confident with respect to the salary after five years. They expect their own salary after five years to be 27 percent higher than the average salary. Male Business and Law students are also fairly confident. Students who study Humanities, contrarily, are relatively pessimistic about their own future earnings.

An obvious concern with this measurement of overconfidence is that it may not solely reflect the individual overconfidence per se. A prospective student who expects to earn more than average others in the same field of study might do so on merit. For example, if a student had a much better final grade in secondary school than the average final grade of students in the same field of study, it is probably not unrealistic to expect that this student will earn an above-average salary in the future. Students who expect to earn more than average might simply have above average abilities. The prospective students might evaluate their own abilities in comparison to others in the same field of study when forming their wage expectations. If all students would be able to realistically asses their own abilities and would base their wage expectations on it, the overconfidence index would just measure relative abilities of students. However, our descriptive evidence as presented above suggest otherwise. Even though female students performed on average better in secondary school, they do have a lower overconfidence on average. Nonetheless, the true abilities of students could have an effect on the overconfidence ratio. There could be further other plausible reasons why a student expects to earn more than average others. Some students might already have good prospects for a well-paying job at the workplace of their parents or other family members after their studies. So, have close relationships to high ranked employees or owners of companies could also play a role when students expect to earn above average salaries. A further possibility why students might expect their future income to differ from the average income might be that they have very strong preferences for non-pecuniary aspects of jobs. A student who values work flexibility, the atmosphere at the workplace, and the enjoyment of a job as more important than the salary might realistically expect to earn less than average others. Therefore, one should be cautious when interpreting the means of our overconfidence index. This index is unlikely to be a perfect measurement of the true overconfidence of the students. The main purpose of this study, however, is not to perfectly measure overconfidence of students, but to evaluate the impact of overconfidence on gender differences in wage expectations. Throughout our estimation procedure, which will be outlined in the next section, we control for the final grade in secondary school of the student, which can be viewed as a proxy for her ability. Moreover, we include dummy variables indicating whether the parents of a student graduated from college, whether they did so in the same field of study for which the student is applying, and whether the student expects to receive financial support ("BAfoeG") while studying. We also include dummy variables indicating answers of the students to questions regarding income importance. Such views are probably closely linked to students' preferences for non-pecuniary aspects of jobs. Thus, above mentioned, potential concerns with regard to our specific overconfidence measure should not pose major problem for the purpose of this study.

# 3 Methodology

## 3.1 Linear Regression

The simplest way to get an estimator of the gender gap in wage expectations that cannot be explained by differences between male and female students in the observed characteristics and to evaluate the impact of overconfidence on wage expectations is to run an OLS regression of the linear model

$$\ln y_i = \delta \cdot d_{Fi} + \boldsymbol{x_i}\boldsymbol{\beta} + v_i. \tag{2}$$

Where the outcome variable, $\ln y_i$, is the logarithm of the own expected starting salary (or salary after five years on the job) of individual $i$, $d_{Fi}$ is a dummy variable which is equal to one if the student is female and $\delta$ is its coefficient. $\boldsymbol{x_i}$ is a $1 \times (K+1)$ vector that contains all the observed characteristics of the students and a constant. $\boldsymbol{\beta}$ is the corresponding $(K+1) \times 1$ coefficient vector of $\boldsymbol{x_i}$ and $v_i$ denotes the error term. By controlling for all the observed variables the overall gender gap in wage expectations is purged of the part that is caused by the different characteristics of males and females. Thus, the estimator of $\delta$ reflects the gender gap in wage expectation that would exist if male and female students would not differ in the observed characteristics.

The coefficient of the overconfidence index can be interpreted as the effect of an increase in overconfidence on the expectation of the own future salary of a student if the overconfidence index is exogenous in equation (2). Unobserved factors are unproblematic with regard to the exogeneity of the overconfidence index as long as they equally affect the expected own salary and the expected average salary, so both parts of the ratio. Only unobserved factors that affect the relative difference between the expected own salary and the expected average salary cause a bias in the estimated coefficient of the overconfidence index. Unobserved factors that might have an effect on the overconfidence index are the ability of the students, the advantage of having already good prospects for a well paying job after graduation, and the preferences of the students regarding non-pecuniary aspects of jobs, as was discussed in the previous section. We control for the final grade in secondary school, the family background, and the view on the importance of income. Thus, these factors should not cause a bias in the estimated impact of the overconfidence index on wage expectations.

Even though a simple linear regression can already show how much of the gender gap in wage expectations can be explained by the observed variables it is not possible to determine how much of the explained part can be attributed to a certain variable. Therefore, we use so-called decomposition methods, described below, for an extensive analysis of the gender difference in wage expectations and to determine the contribution of overconfidence to this difference.

## 3.2 Oaxaca-Blinder Decomposition

Mean differences between two groups in an outcome variable can be analyzed by applying the well known decomposition method proposed by Oaxaca (1973) and Blinder (1973). The method allows to divide the difference in the mean of an outcome variable between two groups into a part that can be explained by different group characteristics and into an unexplained part.

For the standard Oaxaca-Blinder decomposition, it is assumed that the outcome variable can be expressed by a linear model that is separable into observed and unobserved characteristics (Firpo et al., 2011). So we assume that the logarithm of the own expected salary, $\ln y_{gi}$, is given by the following model for males ($M$) and females ($F$), respectively:

$$\ln y_{gi} = \boldsymbol{x_i \beta_g} + v_{gi} \qquad \text{for} \quad g = M, F. \tag{3}$$

Where $\boldsymbol{x_i}$ is again the $1 \times (K+1)$ vector containing the observed explanatory variables and a constant, $\boldsymbol{\beta_g}$ is the corresponding $(K+1) \times 1$ coefficient vector and $v_{gi}$ represents the error term that is unobserved. Furthermore, it is assumed that $\mathbb{E}[v_{gi}|\boldsymbol{x_i}, d_F = d] = 0$ holds, for $d = 0, 1$. Under these assumptions the overall mean difference between females and males, $\Delta_O^\mu$, can be decomposed as follows (Firpo et al., 2011):[7]

$$\begin{aligned}\Delta_O^\mu &= \mathbb{E}[\ln y|d_F = 1] - \mathbb{E}[\ln y|d_F = 0] \\ &= \underbrace{\mathbb{E}[\boldsymbol{x}|d_F = 1] \cdot (\boldsymbol{\beta_F} - \boldsymbol{\beta_M})}_{\Delta_U^\mu} + \underbrace{(\mathbb{E}[\boldsymbol{x}|d_F = 1] - \mathbb{E}[\boldsymbol{x}|d_F = 0]) \cdot \boldsymbol{\beta_M}}_{\Delta_E^\mu}.\end{aligned}$$

So the overall mean difference between females and males in their wage expectation, $\Delta_O^\mu$, can be decomposed into the two components $\Delta_U^\mu$ and $\Delta_E^\mu$. The first component $\Delta_U^\mu$ is the unexplained part of the mean difference and captures the difference between the male and female coefficient vector $\boldsymbol{\beta}$. Hence, it is the part that is caused by the difference in the way females and males form their wage expectations given their personal characteristics $\boldsymbol{x}$. The second component $\Delta_E^\mu$, in contrast, is the explained part, which captures the part of the mean difference that can be explained by the difference between female and male students in their personal characteristics.

The estimation of the two components $\Delta_U^\mu$ and $\Delta_E^\mu$ is straightforward. One can simply replace $\mathbb{E}[\boldsymbol{x}|d_F]$ by the respective sample mean $\bar{\boldsymbol{x}}_{\boldsymbol{g}}$ and plug in the estimator of the respective coefficient vector $\hat{\boldsymbol{\beta}}_{\boldsymbol{g}}$, which can be obtained by running an OLS regression of equation (3) for females and males, respectively (Firpo et al., 2011). The estimated overall mean difference, $\hat{\Delta}_O^\mu$, can thus be decomposed as follows:

$$\hat{\Delta}_O^\mu = \overline{\ln y_F} - \overline{\ln y_M} = \underbrace{\bar{\boldsymbol{x}}_{\boldsymbol{F}}(\hat{\boldsymbol{\beta}}_{\boldsymbol{F}} - \hat{\boldsymbol{\beta}}_{\boldsymbol{M}})}_{\hat{\Delta}_U^\mu} + \underbrace{(\bar{\boldsymbol{x}}_{\boldsymbol{F}} - \bar{\boldsymbol{x}}_{\boldsymbol{M}})\hat{\boldsymbol{\beta}}_{\boldsymbol{M}}}_{\hat{\Delta}_E^\mu}. \tag{4}$$

Due to the additive linearity assumption it is also very easy to compute a detailed decomposition of the overall mean difference, which further divides the two components $\hat{\Delta}_U^\mu$ and $\hat{\Delta}_E^\mu$ into the contributions of the different explanatory variables (Firpo et al., 2011).

One problem with the Oaxaca-Blinder decomposition is that the estimation of the explained and the unexplained part is inconsistent when the linearity assumption is not satisfied (Barsky et al., 2002). In equation (3) it was assumed that the relationship between the logarithm of the expected salary and the explanatory variables is linear. However, this might not necessarily be true. If the relationship is nonlinear the average counterfactual wage expectations that females would have if they would form their expectations the same way as males do is not equal to $\mathbb{E}[\boldsymbol{x}|d_F = 1] \cdot \boldsymbol{\beta_M}$ (Firpo et al., 2011). This problem can be solved by using a reweighted regression approach. The idea of this approach is to use a reweighting function that makes the characteristics of male students similar to those of

[7] In this case the male coefficient vector is chosen to form the counterfactual distribution. Alternatively, the female or pooled coefficient vector can be used (Firpo et al., 2011).

female students.[8] The reweighted regression approach allows to measure the specification error, which occurs when the regression model is misspecified. A specification error close to zero indicates that the the linear model is accurate. In addition, we need to calculate the reweighting error. If the reweighting function is consistently estimated the reweighting error should be close to zero.

### 3.3 RIF Decomposition

To get a better understanding of the gender gap in wage expectations, it can be interesting to not only consider the difference at the mean but also the gender differences at other parts of the wage expectation distribution. The glass ceiling effect, which is often found in the distribution of realized wages, might already be present in the distribution of wage expectations. Furthermore, it can be interesting to see whether the contribution of the gender difference in overconfidence varies along the wage expectation distribution.

To be able to do this the gender gaps at different quantiles of the wage expectations distribution have to be decomposed as in the Oaxaca-Blinder decomposition, described above. Unfortunately, this is a more difficult task than in the case of the mean. It is not possible to decompose quantiles by simply using quantile regressions instead of standard OLS regressions and applying the Oaxaca-Blinder decomposition. The reason for this is that the Oaxaca-Blinder decomposition relies on the law of iterated expectations which does apply in the case of the mean but not in the case of quantiles.

The RIF-regression approach proposed by Firpo et al. (2018), however, allows to perform a detailed decomposition at different quantiles of the wage expectation distribution. The idea of this method is to replace the outcome variable, $\ln y$, by a recentered influence function of the respective quantile, $Q_\tau$, which can be expressed as

$$\mathrm{RIF}(\ln y; Q_\tau) = Q_\tau + \mathrm{IF}(\ln y; Q_\tau) = Q_\tau + \frac{\tau - \mathbb{1}\{\ln y \leq Q_\tau\}}{f_{\ln y}(Q_\tau)}. \tag{5}$$

Where $\mathrm{IF}(\ln y; Q_\tau)$ is the influence function of the quantile, $Q_\tau$, $f_{\ln y}(\cdot)$ is the density of the marginal distribution of $\ln y$, and $\mathbb{1}\{\cdot\}$ is an indicator function taking on the value one if the condition in $\{\cdot\}$ is true (Firpo et al., 2009). Influence functions are an often used tool in the robust statistics literature and represent the contribution of an individual observation to a given distributional statistic. A useful property of influence functions is that by definition $\mathbb{E}[\mathrm{IF}(\ln y; Q_\tau)] = \int \mathrm{IF}(\ln y; Q_\tau) \cdot dF(\ln y) = 0$. Thus, the expected value of the recentered influence function, $\mathrm{RIF}(\ln y; Q_\tau)$, of the quantile $Q_\tau$ is equal to $Q_\tau$ itself (Firpo et al., 2018). Using the law of iterated expectations the unconditional quantile,

[8] It is also possible to choose females as the reference group and use a reweighting function that makes the characteristics of females similar to those of males.

$Q_\tau$, can therefore be written as

$$Q_\tau = \mathbb{E}[\mathbb{E}(\mathrm{RIF}(\ln y; Q_\tau)|\boldsymbol{x})] = \int \mathbb{E}[\mathrm{RIF}(\ln y; Q_\tau)|\boldsymbol{x}] \cdot \boldsymbol{dF_x}(\boldsymbol{x}).$$

For the simplest form of the RIF regression approach it is assumed that the conditional expectation of the recentered influence function can be modelled as a linear function of the explanatory variables $\boldsymbol{x}$ (Firpo et al., 2011):

$$\mathbb{E}[\mathrm{RIF}(\ln y_i; Q_\tau)|\boldsymbol{x_i}] = \boldsymbol{x_i \gamma_\tau} + \upsilon_{\tau i}. \tag{6}$$

Moreover, it is again assumed that the error term, $\upsilon_{\tau i}$, satisfies the condition $\mathbb{E}[\upsilon_{\tau i}|\boldsymbol{x}] = 0$. Under these assumptions the unconditional expectation of $\mathrm{RIF}(\ln y; Q_\tau)$, which is equal to the quantile $Q_\tau$, can be expressed as

$$Q_\tau = \mathbb{E}[\mathrm{RIF}(\ln y; Q_\tau)] = \mathbb{E}[\mathbb{E}(\mathrm{RIF}(\ln y; Q_\tau)|\boldsymbol{x})] = E[\boldsymbol{x}] \cdot \boldsymbol{\gamma_\tau} \tag{7}$$

Consequently the coefficient $\gamma_{\tau j}$, which can easily be estimated by running an OLS regression, can now be interpreted as the effect of an increase in mean value of $x_j$ on the $\tau^{th}$ unconditional quantile of $\ln y$. This means that the gender gap at a certain quantile of the wage expectation distribution can be decomposed as in the standard Oaxaca-Blinder decomposition when $\ln y$ is replaced by the respective $\mathrm{RIF}(\ln y; Q_\tau)$.

Before this can be done though the estimates of $\mathrm{RIF}(\ln y; Q_\tau)$ have to be computed. One can do this by inserting the respective sample quantile, $\hat{Q}_\tau$, and the estimator of the density of the wage expectations at that point, $\hat{f}_{\ln y}(\hat{Q}_\tau)$, which can be obtained by kernel density estimation (Firpo et al., 2009). This has to be done separately for male and female students. When the estimates, $\widehat{\mathrm{RIF}}(\ln y; Q_\tau)$, are computed for both groups it can be proceeded as in the standard Oaxaca-Blinder decomposition.

The explained and the unexplained part are again only consistently estimated when the linearity assumption is true. Since RIF models are unlikely to be linear for distributional statistics besides the mean, it is even more important in case of RIF decompositions to use a reweighting approach and check the specification error (Firpo et al., 2011, 2018).

# 4 Results

## 4.1 Linear Regressions

The estimated overall gender gaps in wage expectations are shown in column 1 of Table 5 for the starting salary, Panel (a), and the salary after five years, Panel (b), respectively. They are obtained by running an OLS regression of the logarithm of the respective ex-

Table 5: Linear Regressions

| | (1) No Controls | (2) Pooled | (3) Male | (4) Female |
|---|---|---|---|---|
| Panel (a): Starting Salary | | | | |
| Female | -0.1717*** | -0.0989*** | | |
| | (0.0247) | (0.0249) | | |
| Overconfidence Start | | 0.2013*** | 0.1677*** | 0.2900*** |
| | | (0.0248) | (0.0293) | (0.0477) |
| Controls | | x | x | x |
| Panel (b): Salary After Five Years | | | | |
| Female | -0.1656*** | -0.1015*** | | |
| | (0.0232) | (0.0236) | | |
| Overconfidence Five | | 0.0527*** | 0.0787*** | 0.0448** |
| | | (0.0152) | (0.0267) | (0.0188) |
| Controls | | x | x | x |
| Observations | 2048 | 2048 | 870 | 1178 |

*Note:* The outcome variable is the logarithm of the respective adjusted gross salary expectation. In columns 2 to 4 we control for the degree, the field of study, the degree with which the student intends to earn first salary, the view on the importance of income, human capital variables, the personal as well as the family background, and the survey year. Overconfidence is measured by the respective overconfidence index. Standard errors are shown in parenthesis. *, ** and *** indicate statistical significance at the 10%-, 5%-, and 1%-level, respectively.

pected salary on a constant and a dummy variable that is equal to one if the student is female. Female students are less optimistic with regard to both salaries. The expected starting salaries of female students are on average 17.2 percent lower than those of male students. Their expectations of their salary after five years on the job are on average 16.6 percent lower. There are substantial gender gaps in the wage expectations that are highly significant and are similar in size for both salaries. In comparison to the results of Reuben et al. (2017), however, we find smaller gender gaps.[9]

Column 2 show the coefficient estimates of equation (2), where the respective wage expectation is additionally regressed on the overconfidence index and the other observed characteristics of the students. The estimated coefficient of the female dummy variable can now be interpreted as the relative difference in the wage expectations between male and female students that would exist in the absence of any gender differences in the observed characteristics. Controlling for the observed characteristics of the students reduces the gender gap in wage expectations considerably. In the case of the starting salary the gap reduces by roughly 7 percentage points to 9.9 percent. The gender gap in the expectations of the salary after five years is now 10.2 percent, which is roughly 6 percentage points

[9] Reuben et al. (2017) find a gender gap of 31 percent in the expected salary at age 30 and a gender gap of 39 percent in the expected salary at age 45.

lower than the overall gender gap. Thus, already simple linear regressions show that more than a third of the gender gaps in wage expectations can be explained by differences in the observed characteristics between male and female students. Nevertheless, there remain sizable unexplained gender gaps in the wage expectations that are still highly significant.

The overconfidence of a student with regard to the respective salary has a positive impact on the wage expectations. The estimated coefficients of the respective overconfidence index are positive and significantly different from zero in both regressions. However, the effect is much larger in the case of the starting salary. A ten percentage points increase in the overconfidence increases on average the expectation of the starting salary by 2 percent holding all other factors constant. An equivalent change in the overconfidence with regard to the salary after five years only leads to an increase of the expected salary after five years by 0.5 percent on average. The size of the overconfidence effect on the wage expectations also differs between male and female students, which can be seen when the model is estimated for males and females separately. A ten percentage points increase in the overconfidence increases the estimated starting salary by 2.9 percent for females compared to only 1.7 percent for males. Contrarily, in the case of the salary after five years the overconfidence effect is somewhat stronger for males than for females.

## 4.2 Decomposition of the Gender Gap

The results of the standard Oaxaca-Blinder decomposition in case of the expected starting salary are shown in Table 6.[10] We use the male coefficient vector to form the counterfactual distribution.[11] As was already shown by the linear regressions the overall relative gender difference in wage expectations is 17.2 percent in case of the starting salary. 8.5 percentage points of this gender gap can be attributed to differences in the observed characteristics between males and females. 8.7 percentage points of the gap, however, remain unexplained and are down to differences in the male and female coefficient vector. This means that almost 50 percent of the gender gap in the expectation of the own starting salary can be explained by the differences in the observed characteristics.

The results of the detailed decomposition show that the difference in overconfidence between male and female students does significantly contribute to the gender gap in the expected own starting salary. 1.3 percentage points of the gender gap can be attributed to the difference in overconfidence. This means that about 8 percent of the gender gap in the expected starting salary is due to the higher overconfidence of male students. This is

[10] From now on we concentrate on the starting salary. The results for the salary after five years are very similar and are available on request.

[11] The decomposition results are very similar when the pooled or female coefficient vector is used as the reference vector instead.

Table 6: Decomposition Results

| | (1) Decomposition Starting Salary | (2) Relative Impact (in %) |
|---|---|---|
| Aggregate Decomposition | | |
| Mean of Males | 8.0467*** | |
| | (0.0191) | |
| Mean of Females | 7.8750*** | |
| | (0.0164) | |
| Gender Gap | 0.1717*** | |
| | (0.0252) | |
| Explained Part | 0.0851*** | 49.54 |
| | (0.0168) | |
| Unexplained Part | 0.0867*** | 50.46 |
| | (0.0272) | |
| Contributions of Covariates to the Explained Part | | |
| Overconfidence Start | 0.0130*** | 7.59 |
| | (0.0045) | |
| Field of Study | 0.0557*** | 32.45 |
| | (0.0143) | |
| Income Importance | 0.0139** | 8.10 |
| | (0.0059) | |
| Other | 0.0024 | 1.39 |
| | (0.0105) | |
| Contributions of Covariates to the Unexplained Part | | |
| Overconfidence Start | -0.1202** | -69.99 |
| | (0.0550) | |
| Field of Study | -0.0274 | -15.95 |
| | (0.0633) | |
| Income Importance | 0.0528 | 30.77 |
| | (0.0439) | |
| Other | -0.0849 | -49.44 |
| | (0.2197) | |
| Constant | 0.2663 | 155.08 |
| | (0.2313) | |
| Observations | 2048 | |

*Note:* The outcome variable is the logarithm of the expected adjusted gross starting salary. The male coefficient vector is used as the reference vector. Overconfidence is measured by the overconfidence index with regard to the starting salary. The standard errors are calculated as described by Jann (2008) and are shown in parenthesis. *, ** and *** indicate statistical significance at the 10%-, 5%-, and 1%-level, respectively. Column 2 shows relative impact on gender gap in percent.

in line with Reuben et al. (2017) who also find that gender differences in overconfidence are partly responsible for the larger expected salaries of male students. The contribution of overconfidence to the gender gap is almost a quarter of the contribution of the field of study. Thus, we find that overconfidence plays an important role for the gender gap in wage expectations. Gender differences in the choice of the field of study together contribute 5.6 percentage points to the gender gap, which corresponds to roughly 32 percent. So, around a third of the gap is down to the fact that males and females sort differently into the fields of study. As was shown in Section 2, females are more likely to apply in a program in Education and Humanities, which are the fields with the lowest expected starting salaries. Males, in contrast, are more likely to apply in a program in Business Studies as well as in a program in Mathematics and Computer Science, which are the fields with the highest expected starting salaries. The third characteristic of the students that significantly contributes to the explained part of the gender gap in the expected starting salaries is the view on the importance of income. 1.4 percentage points of the gap can be attributed to gender differences in the set of dummy variables indicating the answers of the students to the two questions regarding the importance of income. This corresponds to a contribution of around 8 percent in relative terms. A possible explanation for this is that females might be more inclined than males to not just look at the salary but also consider non-pecuniary aspects when searching for a job, because they view their income as less important than males do. This might cause them to already have lower expectations of their own starting salary than males.

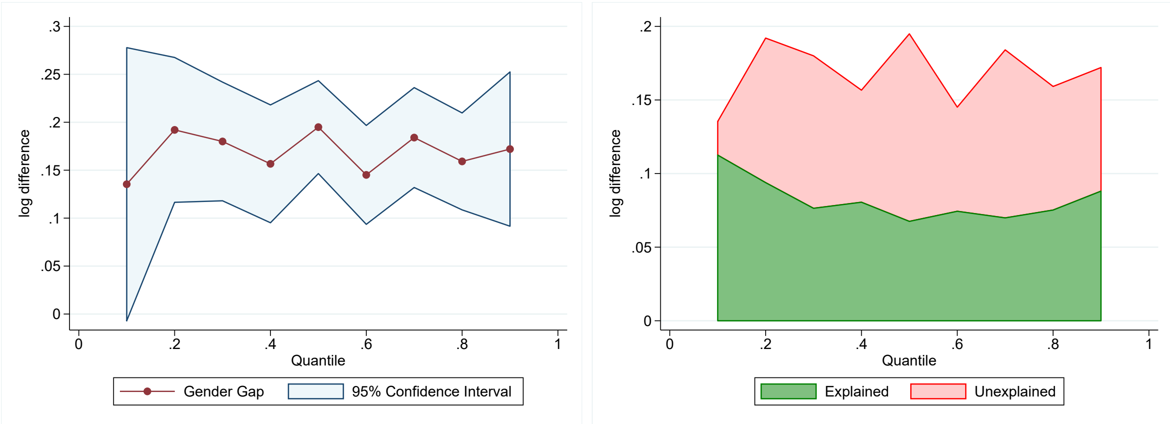
The decomposition of the unexplained part into the contributions of the explanatory variables allows to see how much of the gender gap is due to the difference in the expected return to a certain characteristic. Only the difference in the coefficient of the overconfidence index has an effect on the gender gap that is statistically different from zero. However, the effect is in the opposite direction. It even reduces the gender difference in the expected starting salary. The main contribution to the unexplained part of the gender gap in the expected starting salary comes from the difference in the male and female intercepts, which capture unobserved factors. Hence, it seems not to be the case that males have higher expectations regarding their starting salaries because they expect higher returns to certain characteristics than females do.

## 4.3 Heterogenous Effects

### Heterogeneity along the Wage Expectation Distribution

The gender gap in the wage expectations and the causes of it might differ between different points in the wage expectation distribution. To see whether this is the case we apply the

Figure 1: Aggregate Decomposition: Expected Starting Salary

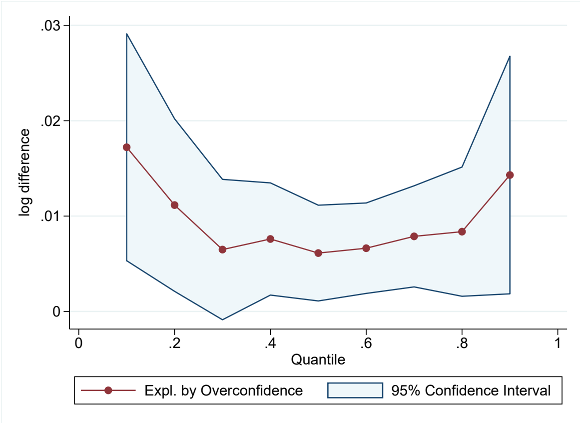



*Note:* Males are chosen as the reference group. Standard errors are bootstrapped (400 replications).

RIF-regression approach, described in Section 3 3.3. The results are depicted in Figures 1 and 2. The gender gap is evaluated at every decile of the wage expectation distribution. Panel (a) of Figure 1 depicts the gender gap in the expected starting salary along the distribution of the expected starting salaries. The gender gap is present at all deciles of the distribution. The gap varies from 13.5 percent at the 10%-quantile to 19.5 percent at the median. As shown above, the mean of the expected starting salaries of males is 17.2 percent higher than the mean of females. So whereas the gender gap at the bottom of the expected starting salary distribution is considerably smaller than at the mean it is larger in the middle of the distribution. So although the size of the gender gap does vary along the wage expectation distribution, we do not find a glass ceiling or sticky floor effect. However, the reason for this could be that our sample only includes prospective college students who are a group of relatively high skilled individuals.

Panel (b) of Figure 1 depicts how much of the gender gap in wage expectations at the different quantiles can be explained by gender differences in the observed characteristics. In contrast to the overall gender gap the explained part of the gender gap varies considerably less along the distributions of the expected starting salary. Hence, the variation of the overall gender gaps is mainly down to the variation of the unexplained part. The explained part of the gap lies between 11.2 and 6.8 percentage points. In comparison, the explained part of the gender gap at the mean is 8.5 percentage points. It is somewhat larger at the edges of the distribution but does not vary a lot in between. At the 10%-quantile the explained part is the largest even though the overall gender gap in the expected starting salary is the smallest at that quantile. Contrarily, the explained part is the smallest at the median of the distribution where the overall gender gap is the largest. This means that despite the low variation of the explained part the share of the explained

Figure 2: Contribution of Overconfidence




*Source:* Survey of prospective university students, own calculations. *Note:* Males are chosen as the reference group. Standard errors are bootstrapped (400 replications).

part at the respective gender gap does vary a lot. Whereas 83 percent of the gender gap in the expected starting salary at the 10%-quantile can be explained by differences in the observed characteristics less than 35 percent of the gap can be explained at the median.

Figure 2 shows the contribution of overconfidence to the explained part of the gender gap in wage expectations along the distribution. Differences between male and female students with regard to the overconfidence do have a significant positive effect on the gender gap in the expected starting salary at every evaluated quantile. The part attributed to gender differences in overconfidence varies, however, between 1.7 and 0.6 percentage points. In line with the total explained part the contribution of the overconfidence is higher at the edges of the distribution than in the middle. The contrast between the edges and the middle of the distribution is even larger in relative terms. Whereas gender differences in overconfidence are responsible for 13 percent of the gap at the 10%-quantile and for 8 percent of the gap at the 90%-quantile, they are only responsible for 3 percent of the gap at the median. The U-shape of the contribution of overconfidence indicates that the female students with the lowest expectations are particularly underconfident and that the male students with the highest expectations are particularly overconfident.

## Heterogeneity between Subgroups of Students

We also analyze whether there are heterogenous effects for some subgroups of the students. The results of the standard Oaxaca-Blinder decomposition for these different subgroups are shown in Tables 7 to 9.

First, in Table 7 we differentiate between prospective Medicine students, who make up roughly a quarter of all students in the sample, and all other prospective students. As was

Table 7: Decomposition Results: Medicine vs. No Medicine

| | (1) Medicine | (2) No Medicine |
|---|---|---|
| Aggregate Decomposition | | |
| Mean of Males | 8.0038*** | 8.0567*** |
| | (0.0563) | (0.0202) |
| Mean of Females | 7.8873*** | 7.8700*** |
| | (0.0323) | (0.0193) |
| Gender Gap | 0.1165* | 0.1867*** |
| | (0.0649) | (0.0280) |
| Explained Part | 0.0420 | 0.0926*** |
| | (0.0627) | (0.0183) |
| Unexplained Part | 0.0746 | 0.0941*** |
| | (0.0816) | (0.0292) |
| Contributions of Covariates to the Explained Part | | |
| Overconfidence Start | 0.0452** | 0.0079* |
| | (0.0204) | (0.0043) |
| Field of Study | | 0.0592*** |
| | | (0.0214) |
| Income Importance | 0.0138 | 0.0079 |
| | (0.0228) | (0.0067) |
| Other | -0.0170 | 0.0176 |
| | (0.0637) | (0.0219) |
| Contributions of Covariates to the Unexplained Part | | |
| Overconfidence Start | -0.1710 | -0.0344 |
| | (0.1069) | (0.0795) |
| Field of Study | | -0.0314 |
| | | (0.0794) |
| Income Importance | 0.0241 | 0.0662 |
| | (0.1181) | (0.0481) |
| Other | -0.2827 | -0.0065 |
| | (0.6911) | (0.2496) |
| Constant | 0.5042 | 0.1003 |
| | (0.7417) | (0.2577) |
| Observations | 506 | 1542 |
| Share of Females | 0.6759 | 0.5422 |

*Note:* The outcome variable is the logarithm of the expected adjusted gross starting salary. In all columns the male coefficient vector is used as the reference vector. Overconfidence is measured by the overconfidence index with regard to the starting salary. The standard errors are calculated as described by Jann (2008) and are shown in parenthesis. *, ** and *** indicate statistical significance at the 10%-, 5%-, and 1%-level, respectively.

highlighted in Section 2, students applying in a program in Medicine are by far the most overconfident subgroup. This raises the question whether the result that gender differences in overconfidence partly explain the gender gap in wage expectations is solely driven by Medicine students. The results show that the overall gender gap in the expected starting salary as well as the size of the explained part is substantially smaller for Medicine students than for the other students. However, the part that is attributed to gender differences in overconfidence is much larger for Medicine students. 4.5 percentage points of the gender gap in the expected starting salary, which corresponds to approximately 39 percent in relative terms, can be explained by the difference in overconfidence for Medicine students. For the other students, in contrast, only 0.8 percentage points corresponding to around 4 percent of the gap are caused by the difference in overconfidence. But the contribution is still weakly significant. This shows that the part of the gender gap in the expected starting salary that can be explained by the larger overconfidence of male students is mainly but not solely driven by Medicine students.

Second, in Table 8 the sample is divided into the group of students applying in a STEM field and the group of students applying in non-STEM fields. In the sample Mathematics and Computer Science as well as Natural Sciences can be classified as STEM fields and all other fields of study as non-STEM fields. Thus, the STEM group comprises 334 students and the non-STEM group 1714 students. Not only are wages in occupations that are related to STEM fields on average higher than in occupations related to non-STEM fields, but it is also expected that STEM employment opportunities will grow in the future. Nonetheless, females remain underrepresented in STEM fields and STEM related occupations (Osikominu and Pfeifer, 2018). Therefore, it might be interesting to see whether there is a difference with regard to the gender gap in wage expectations between students applying in a STEM-field and students applying in a non-STEM field. The gender gap in the expected starting salary is substantially larger in STEM fields than in non-STEM fields. Female students in STEM fields on average expect to earn 29.4 percent less at labor market entry than their male counterparts. The equivalent gender gap in non-STEM fields is only 13.5 percent. Moreover, while a considerable part of the gender gap in non-STEM fields can be explained by differences in the observed characteristics almost the whole gender gap in STEM-fields remains unexplained. So the very large gender gap in STEM fields is not caused by differences in the observed characteristics between male and female students in STEM fields but by unobserved factors. The contribution of overconfidence to the explained part of the gender gap in STEM fields is smaller compared to the full sample and is insignificant. In non-STEM fields the contribution of the differences in overconfidence is a bit larger than in the full sample.

Table 8: Decomposition Results: STEM vs. No STEM

| | (1) STEM | (2) No STEM |
|---|---|---|
| Aggregate Decomposition | | |
| Mean of Males | 8.1633*** | 8.0110*** |
| | (0.0388) | (0.0222) |
| Mean of Females | 7.8691*** | 7.8757*** |
| | (0.0550) | (0.0174) |
| Gender Gap | 0.2942*** | 0.1353*** |
| | (0.0673) | (0.0282) |
| Explained Part | 0.0235 | 0.0696*** |
| | (0.0567) | (0.0179) |
| Unexplained Part | 0.2706*** | 0.0657** |
| | (0.0840) | (0.0295) |
| Contributions of Covariates to the Explained Part | | |
| Overconfidence Start | 0.0059 | 0.0149*** |
| | (0.0085) | (0.0055) |
| Field of Study | 0.0487 | 0.0384*** |
| | (0.0360) | (0.0126) |
| Income Importance | 0.0061 | 0.0136* |
| | (0.0194) | (0.0070) |
| Other | -0.0372 | 0.0028 |
| | (0.0424) | (0.0126) |
| Contributions of Covariates to the Unexplained Part | | |
| Overconfidence Start | -0.3389 | -0.1104** |
| | (0.3271) | (0.0562) |
| discipline | 0.0706 | -0.0683 |
| | (0.0545) | (0.0675) |
| Income Importance | 0.0885 | 0.0394 |
| | (0.1316) | (0.0486) |
| Other | -0.8780 | 0.0459 |
| | (0.8629) | (0.2366) |
| Constant | 1.3284 | 0.1591 |
| | (0.9561) | (0.2459) |
| Observations | 334 | 1714 |
| Share of Females | 0.3892 | 0.6114 |

*Note:* The outcome variable is the logarithm of the expected adjusted gross starting salary. In all columns the male coefficient vector is used as the reference vector. Overconfidence is measured by the overconfidence index with regard to the starting salary. The standard errors are calculated as described by Jann (2008) and are shown in parenthesis. *, ** and *** indicate statistical significance at the 10%-, 5%-, and 1%-level, respectively.

Table 9: Decomposition Results: Freshmen vs. Master Students

| | (1) Freshmen | (2) Master |
|---|---|---|
| Aggregate Decomposition | | |
| Mean of Males | 8.0514*** | 8.0197*** |
| | (0.0204) | (0.0599) |
| Mean of Females | 7.8548*** | 8.0191*** |
| | (0.0179) | (0.0399) |
| Gender Gap | 0.1966*** | 0.0006 |
| | (0.0271) | (0.0720) |
| Explained Part | 0.0730*** | 0.2019*** |
| | (0.0182) | (0.0748) |
| Unexplained Part | 0.1236*** | -0.2013** |
| | (0.0293) | (0.0881) |
| Contributions of Covariates to the Explained Part | | |
| Overconfidence Start | 0.0134*** | 0.0049 |
| | (0.0051) | (0.0108) |
| Field of Study | 0.0444*** | 0.0966** |
| | (0.0153) | (0.0391) |
| Income Importance | 0.0127** | 0.0715* |
| | (0.0059) | (0.0377) |
| Other | 0.0025 | 0.0289 |
| | (0.0115) | (0.0543) |
| Contributions of Covariates to the Unexplained Part | | |
| Overconfidence Start | -0.1046* | -0.2208 |
| | (0.0588) | (0.1740) |
| Field of Study | 0.0058 | -0.0402 |
| | (0.0762) | (0.0729) |
| Income Importance | 0.0550 | 0.1390 |
| | (0.0474) | (0.1272) |
| Other | 0.1163 | -0.9832 |
| | (0.2436) | (0.6809) |
| Constant | 0.0512 | 0.9039 |
| | (0.2556) | (0.7066) |
| Observations | 1776 | 272 |
| Share of Females | 0.5816 | 0.5331 |

*Note:* The outcome variable is the logarithm of the expected adjusted gross starting salary. In all columns the male coefficient vector is used as the reference vector. Overconfidence is measured by the overconfidence index with regard to the starting salary. The standard errors are calculated as described by Jann (2008) and are shown in parenthesis. *, ** and *** indicate statistical significance at the 10%-, 5%-, and 1%-level, respectively.

Finally, in Table 9 we also compare students applying for a Bachelor program or a State Examination program with students applying for a Master program. The subgroup of Master applicants is much smaller but does include prospective students of all fields of study apart from Law. Applicants for a Bachelor and a State Examination program are usually undergraduates, who have not attended college before and are often referred to as freshmen. Applicants for a Master program, in contrast, have already completed a Bachelor degree. Several studies show that students increase their knowledge of average salaries in the labor market during their time in college (e.g. Arcidiacono et al., 2017; Betts, 1996; Botelho and Pinto, 2004). Moreover, Klößner and Pfeifer (2018), who use the same data set, find that female students make slightly larger estimation errors when predicting actual salaries, which indicates that female applicants are less well informed about actual salaries than male applicants. Female students tend to underestimate average salaries to a larger extend than male students. Thus, the gender difference in wage expectations might be in part due to the larger estimation errors of females. The gender difference in the information about actual salaries, however, might mitigate or even disappear over the course of college, when students update their beliefs about salaries in their field of study. The gender gap in the expected starting salary for freshmen is a bit larger compared to the full sample. Interestingly, there is no gender difference at all in the estimated starting salary for students applying for a Master program. However, the reason for this is not that male and female applicants for a Master program do not differ in the observed characteristics. Quite in the contrary, the explained part is more than two times as large as in the full sample. But, the explained part is completely offset by a negative unexplained part. A possible explanation for this is that prospective Master students, in contrast to freshmen, base their own wage expectations mainly on the observed average salaries in the labor market rather than on their own characteristics. But, one should be cautious interpreting this result since the number of prospective Master students in the sample is relatively small. The contribution of overconfidence to the explained part is smaller than in the full sample and is statistically insignificant.

# 5 Robustness Checks

As discussed in Section 3, the decomposition results are only valid if the linearity assumption is satisfied. Therefore, we also apply the reweighted-regression approach, described above. In case of the standard Oaxaca-Blinder decomposition we find a specification error of 0.002 which is statistically not different from zero.[12] Moreover, the reweighting error is also very close to zero and statistically insignificant, which indicates that the reweighting

[12] All other results of the decompositions using the reweighting approach are available on request.

factors are consistently estimated. In case of the RIF decompositions we also find no statistically significant specification error at any point of the distribution of the expected starting salary. This suggests that the decomposition results without reweighting are not invalid because of a misspecification of the regression models. Hence, we conclude that the linear specification is quite accurate and continue with it.

As outlined in Section 2, the mean of the overconfidence index regarding the expected starting salary is larger than one, even though more students are underconfident than overconfident. This indicates that there are a few students who are extremely overconfident. Hence, it could be the case that our results are mainly driven by such outliers. Furthermore, there are also some concerns that the overconfidence index might be endogenous. In Section 3 it was discussed that the fact that the dependent variable is part of the overconfidence index might cause a bias in the estimated coefficient. To check the robustness of our results we use other overconfidence measures that are likely to be less affected by these concerns. Table 10 shows the results of the Oaxaca-Blinder decompositions using these different overconfidence measures.

First, we use a dummy variable that is equal to one if the student expects to earn more than average others instead of the overconfidence index. The results are shown in column 1. The decomposition results using the dummy indicator of overconfidence are less prone to outliers, since the dummy indicator is not affected by the level of overconfidence. The dummy indicator does not show differences between students who expect to earn slightly more than average others and students who expect to earn much more than average others. For the same reason it is probably also less likely that the estimated coefficient of the dummy indicator of overconfidence is biased. The decomposition result using the dummy indicator is very similar to our previous result using the overconfidence index. The contribution of overconfidence to the explained part is even a bit larger.

In column 2 we use again a dummy indicator of overconfidence, but this time with regard to the salary after five years. That way the overconfidence measure is even less prone to an endogeneity problem since the outcome variable is no longer part of the overconfidence measure. The contribution of overconfidence to the explained part of the gender gap in the expected starting salary is then again larger compared to the result when the overconfidence index is used.

Lastly, in column 3 a dummy variable that indicates whether the student is overconfident with regard to any of the two salaries is used. The dummy variable is equal to one if the student expects to earn more than average others at labor market entry or after five years on the job or at both points. The decomposition result does not differ much from the result when the overconfidence index is used. As in the two previous cases,

Table 10: Decomposition Results With Other Overconfidence Measures

| | (1) $oc_{start} > 1$ | (2) $oc_{five} > 1$ | (3) $oc_{start} > 1$ $\vee$ $oc_{five} > 1$ |
|---|---|---|---|
| **Aggregate Decomposition** | | | |
| Mean of Males | 8.0467*** | 8.0467*** | 8.0467*** |
| | (0.0191) | (0.0191) | (0.0191) |
| Mean of Females | 7.8750*** | 7.8750*** | 7.8750*** |
| | (0.0164) | (0.0164) | (0.0164) |
| Gender Gap | 0.1717*** | 0.1717*** | 0.1717*** |
| | (0.0252) | (0.0252) | (0.0252) |
| Explained Part | 0.0868*** | 0.0921*** | 0.0910*** |
| | (0.0170) | (0.0172) | (0.0171) |
| Unexplained Part | 0.0849*** | 0.0796*** | 0.0807*** |
| | (0.0273) | (0.0277) | (0.0277) |
| **Contributions of Covariates to the Explained Part** | | | |
| Overconfidence | 0.0172*** | 0.0194*** | 0.0188*** |
| | (0.0053) | (0.0058) | (0.0056) |
| Field of study | 0.0519*** | 0.0552*** | 0.0544*** |
| | (0.0143) | (0.0145) | (0.0145) |
| Income Importance | 0.0121** | 0.0120** | 0.0120** |
| | (0.0059) | (0.0060) | (0.0060) |
| Other | 0.0057 | 0.0056 | 0.0058 |
| | (0.0105) | (0.0106) | (0.0106) |
| **Contributions of Covariates to the Unexplained Part** | | | |
| Overconfidence | 0.0027 | -0.0042 | 0.0016 |
| | (0.0098) | (0.0112) | (0.0123) |
| Field of study | -0.0243 | -0.0327 | -0.0295 |
| | (0.0635) | (0.0638) | (0.0638) |
| Income Importance | 0.0550 | 0.0580 | 0.0591 |
| | (0.0440) | (0.0444) | (0.0444) |
| Other | -0.0811 | -0.1015 | -0.0920 |
| | (0.2202) | (0.2216) | (0.2217) |
| Constant | 0.1327 | 0.1600 | 0.1415 |
| | (0.2233) | (0.2245) | (0.2248) |
| Observations | 2048 | 2048 | 2048 |

*Note:* The outcome variable is the logarithm of the expected adjusted gross starting salary. In all columns the male coefficient vector is used as the reference vector. Overconfidence is measured by the respective dummy variable. The standard errors are calculated as described by Jann (2008) and are shown in parenthesis. *, ** and *** indicate statistical significance at the 10%-, 5%-, and 1%-level, respectively.

the contribution of overconfidence to the explained part is a bit larger than when the overconfidence index is used.

Overall, one can conclude that the decomposition results are robust to the use of a different overconfidence measure. The estimated contribution of overconfidence to the explained part of the gender gap in the expected starting salary is always similar in size no matter which of the different overconfidence measures are used. Moreover, neither the size of the explained part nor the contributions of the other observed characteristics changes dramatically when a different overconfidence measure is used. This indicates that the previous results are not only driven by a few outliers. Furthermore, the concern that the estimated coefficient of the overconfidence index is biased upwards seems to be not warranted. If that would be the case the contribution of overconfidence to the explained part should be smaller when these other overconfidence measures are used and not larger.

# 6 Conclusion

We use large and rich survey data on students' beliefs about their future salaries, which is representative for the German student population. We show that female students expect to earn around 17 percent less at labor market entry as well as after five years on the job compared to male students. This finding is relevant for an explanation of the gender wage gap, since previous studies indicate that wage expectations do affect realized wages (Webbink and Hartog, 2004). We further evaluate effect heterogeneity with respect to two dimensions—across different levels of wage expectations as well as across different subgroups. Our analysis of the gender gap along the wage expectation distribution suggests that there is neither a sticky floor nor a glass ceiling effect in the earnings expectations of students. However, the gender gap in wage expectations does differ between some subgroups of students. In STEM fields, for instance, the gender gap is even larger than in the full sample. Contrarily, we do not find a gender difference in wage expectations of prospective Master students.

Furthermore, we measure overconfidence of students by comparing expectations of their own salaries with their expectations of their fellow students' average salaries. As is suggested by previous studies, we find that male students are significantly more overconfident than female students (Barber and Odean, 2001; Bengtsson et al., 2005). Linear regressions of wage expectations on our overconfidence measure show that the overconfidence of students indeed affects their wage expectations. Students who are more overconfident tend to expect higher future salaries. Our decomposition results suggest that a substantial part of the gender difference in wage expectation is down to larger overconfidence of male students. Gender differences in overconfidence seem to be particularly influential at the bottom and at the top of the wage expectation distribution, as results of our RIF decompositions indicate. This raises the question whether overconfidence is an intrinsic characteristic of students that is given by birth or whether it is possible to influence the confidence of individuals. If confidence of individuals can indeed be influenced, policy measures aimed at raising the confidence of females before they leave secondary school might reduce the gender difference in pay.

Moreover, we also find that gender differences in college major choice and in the view on the importance of income significantly contribute to the gender gap in wage expectations. Females seem to choose fields with lower income expectations and seem to have stronger preferences for non-pecuniary aspects of jobs than males. Nonetheless, a large part of the gender gap in wage expectations cannot be explained by the observed characteristics of students. Even in the hypothetical case that female students would not differ from male students in observed characteristics, more than half of the gender gap would still remain. Hence, there must exist other factors that also cause females to have lower wage expectations than males. Nevertheless, by considering a measure of overconfidence in our analysis, we can make new contributions to the existing literature on wage expectations and gender differences.

# References


Arcidiacono, P. (2004). Ability sorting and the returns to college major. *Journal of Econometrics*, 121:343–375.

Arcidiacono, P., Hotz, V. J., Maurel, A., and Romano, T. (2017). Ex Ante Returns and Occupational Choice. *Working Paper*.

Baker, R., Bettinger, E., Jacob, B., and Marinescu, I. (2018). The Effect of Labor Market Information on Community College Students' Major Choice. *Economics of Education Review*, 65:18–30.

Barber, B. M. and Odean, T. (2001). Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investment. *The Quarterly Journal of Economics*, 116(1):261–292.

Barsky, R., Bound, J., Charles, K. K., and Lupton, J. P. (2002). Accounting for the Black–White Wealth Gap: A Nonparametric Approach. *Journal of the American Statistical Association*, 97(459):663–673.

Bengtsson, C., Persson, M., and Willenhag, P. (2005). Gender and Overconfidence. *Economics Letters*, 86(2):199 – 203.

Betts, J. R. (1996). What Do Students Know About Wages? Evidence From a Survey of Undergraduates. *The Journal of Human Resources*, 31(1):27–56.

Blau, F. D. and Kahn, L. M. (2017). The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature*, 55(3):789–865.

Blinder, A. S. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *The Journal of Human Resources*, 8(4):436–455.

Botelho, A. and Pinto, L. C. (2004). Students' Expectations of the Economic Returns to College Education: Results of a Controlled Experiment. *Economics of Education Review*, 23(6):645–653.

Christofides, L. N., Polycarpou, A., and Vrachimis, K. (2013). Gender Wage Gaps, 'Sticky Floors' and 'Glass Ceilings' in Europe. *Labour Economics*, 21:86–102.

Croson, R. and Gneezy, U. (2009). Gender Differences in Preferences. *Journal of Economic Literature*, 47(2):448–474.

Dahlbom, L., Jakobsson, A., Jakobsson, N., and Kotsadam, A. (2011). Gender and Overconfidence: Are Girls Really Overconfident? *Applied Economics Letters*, 18(4):325–327.

Firpo, S. P., Fortin, N. M., and Lemieux, T. (2009). Unconditional Quantile Regressions. *Econometrica*, 77(3):953–973.

Firpo, S. P., Fortin, N. M., and Lemieux, T. (2011). Chapter 1 - Decomposition Methods in Economics. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 4. Elsevier.

Firpo, S. P., Fortin, N. M., and Lemieux, T. (2018). Decomposing Wage Distributions Using Recentered Influence Function Regressions. *Econometrics*, 6.

Hardies, K., Breesch, D., and Branson, J. (2013). Gender differences in Overconfidence and Risk Taking: Do Self-Selection and Socialization Matter? *Economics Letters*, 118(3):442–444.

Jacob, B. A. and Wilder, T. (2011). Educational Expectations and Attainment. In Duncan, G. and Murnane, R., editors, *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*, page 133–162. New York: Russell Sage Press.

Jann, B. (2008). The Blinder–Oaxaca Decomposition for Linear Regression Models. *Stata Journal*, 8(4):453–479.

Klößner, S. and Pfeifer, G. (2018). The Importance of Tax Adjustments when Evaluating Wage Expectations. *The Scandinavian Journal of Economics*.

Nekby, L., Thoursie, P. S., and Vahtrik, L. (2008). Gender and Self-Selection into a Competitive Environment: Are Women More Overconfident than Men? *Economics Letters*, 100(3):405–407.

Niederle, M. and Vesterlund, L. (2007). Do Women Shy Away From Competition? Do Men Compete Too Much? *The Quarterly Journal of Economics*, 122(3):1067–1101.

Nyhus, E. K. and Pons, E. (2012). Personality and the Gender Wage Gap. *Applied Economics*, 44(1):105–118.

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14(3):693–709.

Osikominu, A. and Pfeifer, G. (2018). Perceived Wages and the Gender Gap in STEM Fields. *CEPR Discussion Paper No. DP12719*.

Reuben, E., Wiswall, M., and Zafar, B. (2017). Preferences and Biases in Educational Choices and Labour Market Expectations: Shrinking the Black Box of Gender. *The Economic Journal*, 127(604):2153–2186.

Soll, J. B. and Klayman, J. (2004). Overconfidence in Interval Estimates. *Journal of Experimental Psychology*, 30(2):299–314.

Webbink, D. and Hartog, J. (2004). Can Students Predict Starting Salaries? Yes! *Economics of Education Review*, 23(2):103–113.

Weichselbaumer, D. and Winter-Ebmer, R. (2005). A Meta-Analysis of the International Gender Wage Gap. *Journal of Economic Surveys*, 19(3):479–511.

Wiswall, M. and Zafar, B. (2015). Determinants of College Major Choice: Identification Using an Information Experiment. *The Review of Economic Studies*, 82(2):791–824.

Wiswall, M. and Zafar, B. (2018). Preference for the Workplace, Investment in Human Capital, and Gender. *The Quarterly Journal of Economics*, 133(1):457–507.

Zafar, B. (2013). College Major Choice and the Gender Gap. *The Journal of Human Resources*, 48(3):545–595.

# Appendix

Table 11: Wage Expectations by Field of Study

| | Starting Salary | | Salary after Five Years | |
|---|---|---|---|---|
| Field of Study | Own | Average | Own | Average |
| Business Studies | 3469.84 | 3551.82 | 5279.78 | 5265.16 |
| | (1567.67) | (1589.03) | (2655.85) | (2831.53) |
| Education | 2868.58 | 2977.35 | 3930.69 | 3994.81 |
| | (1214.24) | (1325.89) | (1463.38) | (1548.45) |
| Humanities | 2555.70 | 2684.10 | 3926.60 | 4031.60 |
| | (1285.51) | (1318.76) | (2208.47) | (2303.41) |
| Law Studies | 3427.33 | 3641.43 | 5809.19 | 5802.83 |
| | (2085.11) | (2051.45) | (4044.67) | (3695.47) |
| Math./Comp. Science | 3765.32 | 3874.09 | 5655.97 | 5639.75 |
| | (1729.89) | (1843.99) | (2569.75) | (2630.20) |
| Medicine | 3234.32 | 3246.53 | 5119.19 | 5020.03 |
| | (1826.35) | (1759.70) | (3218.58) | (3315.08) |
| Natural Sciences | 3370.52 | 3419.75 | 5002.12 | 4897.88 |
| | (2061.14) | (1887.75) | (2767.50) | (2631.06) |

*Note:* Table shows means and standard deviations (in parenthesis). Wage expectations are expected gross salaries that were adjusted for misconceptions of the progressive income tax by Klößner and Pfeifer (2018).

Table 12: RIF Decomposition Results: Starting Salary

| | (1)<br>10% | (2)<br>20% | (3)<br>30% | (4)<br>40% | (5)<br>50% |
|---|---|---|---|---|---|
| | | | Aggregate Decomposition | | |
| Quantile of Males | 7.3511*** | 7.6853*** | 7.8590*** | 7.9794*** | 8.1394*** |
| | (0.0506) | (0.0328) | (0.0204) | (0.0231) | (0.0155) |
| Quantile of Females | 7.2157*** | 7.4933*** | 7.6790*** | 7.8228*** | 7.9445*** |
| | (0.0544) | (0.0173) | (0.0236) | (0.0208) | (0.0194) |
| Gender Gap | 0.1354* | 0.1920*** | 0.1800*** | 0.1566*** | 0.1949*** |
| | (0.0727) | (0.0385) | (0.0316) | (0.0313) | (.0247) |
| Explained Part | 0.1126*** | 0.0938*** | 0.0764*** | 0.0805*** | 0.0675*** |
| | (0.0415) | (0.0273) | (0.0220) | (0.0208) | (0.0184) |
| Unexplained Part | 0.0228 | 0.0982** | 0.1036*** | 0.0761** | 0.1274*** |
| | (0.0809) | (0.0470) | (0.0364) | (0.0327) | (0.0293) |
| | | Contributions of Covariates to the Explained Part | | | |
| Overconfidence Start | 0.0172*** | 0.0112** | 0.0065* | 0.0076** | 0.0061** |
| | (0.0061) | (0.0046) | (0.0038) | (0.0030) | (0.0026) |
| Field of Study | 0.0800** | 0.0714*** | 0.0568*** | 0.0647*** | 0.0597*** |
| | (0.0349) | (0.0238) | (0.0181) | (0.0187) | (0.0161) |
| Income Importance | 0.0202 | 0.0072 | 0.0073 | 0.0050 | 0.0035 |
| | (0.0154) | (0.0102) | (0.0071) | (0.0073) | (0.0067) |
| Other | -0.0049 | 0.0041 | 0.0058 | 0.0033 | -0.0018 |
| | (0.0261) | (0.0167) | (0.0138) | (0.0124) | (0.0115) |
| | | Contributions of Covariates to the Unexplained Part | | | |
| Overconfidence Start | -0.1981 | -0.1637 | -0.2120** | -0.2034*** | -0.1336* |
| | (0.1689) | (0.1077) | (0.0827) | (0.0762) | (0.0721) |
| Discipline | -0.0220 | -0.1046 | -0.0403 | -0.0610 | -0.0259 |
| | (0.1589) | (0.0979) | (0.0783) | (0.0753) | (0.0703) |
| Income Importance | -0.0198 | -0.0548 | 0.0026 | -0.0309 | -0.0413 |
| | (0.1135) | (0.0702) | (0.0599) | (0.0541) | (0.0458) |
| Other | 0.5616 | 0.1568 | 0.1427 | 0.0381 | -0.0096 |
| | (0.6886) | (0.3442) | (0.3022) | (0.2832) | (0.2539) |
| Constant | -0.2989 | 0.2645 | 0.2106 | 0.3333 | 0.3378 |
| | (0.7211) | (0.3765) | (0.3187) | (0.2946) | (0.2740) |
| Observations | 2048 | 2048 | 2048 | 2048 | 2048 |

*Note:* The outcome variable is the respective estimated RIF of the logarithm of the adjusted expected gross starting salary. Males are chosen as the reference group. Overconfidence is measured by the overconfidence index with regard to the starting salary. The standard errors are bootstrapped (400 replications) and are shown in parenthesis. *, ** and *** indicate statistical significance at the 10%-, 5%-, and 1%-level, respectively.

Table 13: RIF Decomposition Results: Starting Salary *(continued)*

| | (6) 60% | (7) 70% | (8) 80% | (9) 90% |
|---|---|---|---|---|
| **Aggregate Decomposition** | | | | |
| Quantile of Males | 8.2208*** | 8.3407*** | 8.4574*** | 8.6553*** |
| | (0.0179) | (0.0212) | (0.0170) | (0.0319) |
| Quantile of Females | 8.0756*** | 8.1567*** | 8.2982*** | 8.4832*** |
| | (0.0201) | (0.0183) | (0.0212) | (0.0269) |
| Gender Gap | 0.1451*** | 0.1840*** | 0.1592*** | 0.1720*** |
| | (0.0263) | (0.0266) | (0.0258) | (0.0411) |
| Explained Part | 0.0744*** | 0.0699*** | 0.0752*** | 0.0881*** |
| | (0.0176) | (0.0179) | (0.0191) | (0.0253) |
| Unexplained Part | 0.0707** | 0.1141*** | 0.0840*** | 0.0840** |
| | (0.0304) | (0.0307) | (0.0282) | (0.0421) |
| **Contributions of Covariates to the Explained Part** | | | | |
| Overconfidence Start | 0.0066*** | 0.0079*** | 0.0084** | 0.0143** |
| | (0.0024) | (0.0027) | (0.0035) | (0.0064) |
| Field of Study | 0.0604*** | 0.0534*** | 0.0531*** | 0.0401** |
| | (0.0141) | (0.0143) | (0.0144) | (0.0175) |
| Income Importance | 0.0082 | 0.0106* | 0.0133* | 0.0200** |
| | (0.0067) | (0.0061) | (0.0068) | (0.0091) |
| Other | -0.0008 | -0.0020 | 0.0005 | 0.0137 |
| | (0.0115) | (0.0118) | (0.0119) | (0.0175) |
| **Contributions of Covariates to the Unexplained Part** | | | | |
| Overconfidence Start | -0.1023 | -0.1030 | -0.1323* | -0.1123 |
| | (0.0742) | (0.0743) | (0.0761) | (0.1137) |
| Field of Study | -0.0547 | -0.0515 | -0.0661 | 0.0839 |
| | (0.0659) | (0.0617) | (0.0688) | (0.0854) |
| Income Importance | 0.0251 | 0.0534 | 0.0716 | 0.1198 |
| | (0.0431) | (0.0457) | (0.0496) | (0.0746) |
| Other | 0.0010 | 0.0635 | 0.1805 | -0.2478 |
| | (0.2258) | (0.2329) | (0.2623) | (0.3414) |
| Constant | 0.2016 | 0.1518 | 0.0301 | 0.2404 |
| | (0.2394) | (0.2406) | (0.2659) | (0.3732) |
| Observations | 2048 | 2048 | 2048 | 2048 |

*Note:* The outcome variable is the respective estimated RIF of the logarithm of the adjusted expected gross starting salary. Males are chosen as the reference group. Overconfidence is measured by the overconfidence index with regard to the starting salary. The standard errors are bootstrapped (400 replications) and are shown in parenthesis. *, ** and *** indicate statistical significance at the 10%-, 5%-, and 1%-level, respectively.