I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13517

Overconfidence and Gender Differences in Wage Expectations

Stephanie Briel
Aderonke Osikominu
Gregor Pfeifer
Mirjam Reutter
Sascha Satlukal

JULY 2020

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13517

Overconfidence and Gender Differences in Wage Expectations

Stephanie Briel
University of Hohenheim

Aderonke Osikominu
University of Hohenheim, CEPR, CESifo and IZA

Gregor Pfeifer
University College London, University of Hohenheim, CESifo and IZA

Mirjam Reutter
University of Hohenheim

Sascha Satlukal
University of Hohenheim

JULY 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Overconfidence and Gender Differences in Wage Expectations*

We analyze the impact of (over-)confidence on gender differences in expected starting salaries using elicited beliefs of prospective university students in Germany. According to our results, female students have lower wage expectations and are less overconfident than their male counterparts. Oaxaca-Blinder decompositions of the mean show that 7.7% of the gender gap in wage expectations is attributable to a higher overconfidence of males. Decompositions of the unconditional quantiles of expected salaries suggest that the contribution of gender differences in confidence to the gender gap is particularly strong at the bottom and top of the wage expectation distribution.

JEL Classification: J16, D84, D91, C21

Keywords: gender pay gap, wage expectations, overconfidence, decomposition analyses, unconditional quantile regressions (RIF-Regressions)

Corresponding author:
Sascha Satlukal
University of Hohenheim
Department of Economics
Schloss Hohenheim 1
70599 Stuttgart
Germany
E-mail: sascha.satlukal@uni-hohenheim.de

* The authors are grateful for valuable remarks from Martin Biewen, Bernd Fitzenberger, and Anthony Strittmatter as well as seminar and conference participants at the University of Tübingen, the University of Hohenheim, the Journées LAGV, the International Workshop on Applied Economics of Education, the Econometrics Conference on "Economic Applications of Quantile Regressions 2.0". "Heterogenität von Erträgen und Kosten der Ausbildung in MINT-Berufen" is part of the research program "Netzwerk Bildungsforschung" (Educational Research Network) of the Baden-Württemberg Stiftung.

1 Introduction

Pay differences between men and women are one of the most intensively investigated phenomena in Economics. Although the gender pay gap has been declining over the last decades in developed countries, women still earn considerably less than men (Blau and Kahn, 2017). Gender differences do not only exist in realized wages, but they already emerge in expectations of young people about their future salaries (e.g. Filippin and Ichino, 2005; Kiessling et al., 2019; Reuben et al., 2017). Hence, one reason why women do not earn as much as men might be lower wage expectations as individuals tend to accept jobs that match their beliefs (Reuben et al., 2017). Gender differences in expected salaries in turn may be attributable to differences in confidence of men and women. In fact, men tend to be more confident than women about their abilities (e.g. Bengtsson et al., 2005).

In this paper, we analyze the impact of (over-)confidence on gender differences in the salaries prospective university students expect to earn after graduation. Our empirical analysis makes use of salary expectations elicited in a survey conducted at Saarland University, Germany. During the application process in 2011 and 2012, prospective students were asked to report expectations of their own future salary as well as of the average future salary of other students in the same program. We use this information to construct various measures of confidence by relating the expected salary for oneself to that expected for the average among students in the same field of study and with identical intended degree. Our benchmark confidence index just uses the ratio of the two expectations and takes on a value larger than one if the student expects to earn more than average, which we interpret as overconfidence. We also consider alternative indicators for over- and underconfidence that relate the benchmark index to the rank of a respondent's GPA in secondary school. The alternative confidence indicators address the potential concern that the benchmark confidence index may reflect differences in ability rather than confidence per se. We then use the confidence measures to quantify the contribution of (over-)confidence to gender differences in expected starting salaries. For this purpose, we apply both Oaxaca-Blinder

decompositions at the mean of expected starting salaries as well as decompositions of the unconditional quantiles of expected starting salaries based on the method proposed by Firpo et al. (2009, 2018).

Our results indicate that the average expected starting salary of women is 17 percent lower than that of men. At the same time, men tend to be more overconfident than women. According to our benchmark confidence index, the gender gap in confidence equals almost eight percent. Moreover, confidence has a positive effect on the expected starting salary. These results indicate already that differences in confidence contribute partly to the observed gap in male and female expected starting salaries. Going into details, the results of the Oaxaca-Blinder decomposition at the mean show that 7.7 percent of the overall gender gap (and 15 percent of the explained gap) in expected starting salaries is attributable to the fact that male students are more confident than females. At the same time, gender differences in the wage structure effect of confidence contribute negatively to the overall gender gap suggesting that a unit increase in the mean of confidence increases the mean of the female expected starting salary more strongly than the male one. Alternative decompositions that include the GPA adjusted indicators for over- and underconfidence instead of the benchmark confidence index reveal that the contribution of gender differences in confidence to the explained gap is mainly driven by gender differences in overconfidence as opposed to underconfidence. Furthermore, the decompositions at the unconditional deciles of male and female expected starting salaries show that, while the gender gap is present at all deciles, there is no evidence for sticky floor (wider gap at the bottom) nor glass ceiling effects (wider gap at the top). The contribution of the confidence index to the explained gap, in contrast, exhibits a U-shaped pattern along the deciles. This suggests that gender differences in confidence do not contribute to the gender gap in the dispersion of expected starting salaries as measured by the difference between the ninth and first decile. Finally, our results show that differential sorting of men and women into fields of study accounts for 30 percent of the overall gap (60 percent of the explained gap) at the mean and for similarly large shares of the gaps at the deciles of expected starting

salaries. This pattern parallels the findings in the literature on realized wages that generally attributes the most important share of the explained gap to differential sorting into occupations and industries.

The main innovation of our study is to propose new measures of (over-)confidence and to use them to examine the role of (over-)confidence in shaping the gender gap in wage expectations of prospective university students. Our paper thus contributes to three research areas: gender differences in pay, wage expectations, and overconfidence.

Our study speaks to the literature analyzing the gender gap in actual pay (see e.g. the meta-analysis by Weichselbaumer and Winter-Ebmer (2005) for an overview). Many studies in this line of research document a U-shaped pattern of the gender pay gap along the conditional and unconditional distributions of earnings, meaning that the gender pay gap is particularly pronounced in the lower (sticky floor effect) and upper tail (glass ceiling effect) of wage distributions (Antonczyk et al., 2010; Christofides et al., 2013; Collischon, 2019). Moreover, the largest part of the gender pay gap is typically attributed to the fact that men and women sort differently into occupations and industries. Francesconi and Parey (2018), who focus on university graduates in Germany, a group closely related to our sample, document a gender gap of 20 percent in actual starting salaries. Focusing on prospective students, we are able to document that the gender pay gap is already present at an earlier point in the careers of young people: when prospective students form their expectations about their future salary. Similar to the literature on the gender gap in actual pay we also observe that the differential sorting of men and women into fields of study accounts for the largest share of the gender gap in expected salaries. In contrast to the literature on actual pay, we do not detect sticky floor or glass ceiling effects.

In contrast to the gender gap in actual earnings, gender differences in earnings expectations have been much less under study so far. This might in part be related to the historical reluctance of economists to work with data on subjective expectations, stemming from the doubt whether elicited subjective expectations of survey participants reflect their true expectations. Botelho and Pinto (2004) show that these concerns are unwar-

ranted, and the use of subjective expectations data in economics in general as well as in the context of education and labor economics has increased over the last years (see, e.g., Arcidiacono et al., 2020; Baker et al., 2018; Jacob and Wilder, 2011; Reuben et al., 2017; Zafar, 2013). Specifically, Kiessling et al. (2019) who analyze wage expectations of advanced students in Germany find significant gender gaps in expected starting salaries and in expected lifetime earnings. Moreover, their results suggest that a substantial part of the gap can be attributed to gender differences in sorting into fields of study as well as personal style in pay negotiations. Unlike Kiessling et al. (2019), we focus on the gender gap in the expected starting salaries of prospective students, young individuals who may already have some work experience in their desired branch of business but who are yet to enter university. While Kiessling et al. (2019) consider a range of potential determinants of the gender pay gap in wage expectations they do not study specifically the role of (over-)confidence as we do.

In fact, how confident a student is about their own abilities constitutes an important factor determining expected salary. The previous literature on overconfidence documents that people are often overconfident about their own abilities (Croson and Gneezy, 2009). Moreover, many studies in this strand of the literature find important differences between men and women with regard to overconfidence in various contexts (Barber and Odean, 2001; Bengtsson et al., 2005; Soll and Klayman, 2004; Niederle and Vesterlund, 2007; Dahlbom et al., 2011). Some studies focus on highly specialized, selective, or male-dominated settings which makes it difficult to assess the external validity of the results (Nekby et al., 2008; Hardies et al., 2013). All in all, while most of the literature suggests that men are more overconfident than women, it seems like gender differences in overconfidence depend on the particular setting at hand. Specifically, by linking stereotypes and overconfidence, Bordalo et al. (2019) show that males are more likely to be overconfident in their abilities in domains in which males are said to be good at, whereas females are more likely to be overconfident in domains in which females are said to be good at. In our study, we therefore propose measures of (over-)confidence that are directly related to

an individual's comparative wage expectations for their chosen field of study and degree.

Closest to our study is the work of Reuben et al. (2017) who analyze the impact of overconfidence as well as competitiveness and risk taking on earnings expectations and college major choice of undergraduate students at New York University. Combining experimental and survey data, they show that overconfidence and competitiveness have a positive impact on expected future salaries. Further, they find that male students have considerably higher earnings expectations and that overconfidence as well as competitiveness are responsible for a substantial part of the gender gap. In contrast to Reuben et al. (2017), we go beyond linear regressions and analyze the impact of confidence in more detail applying decomposition techniques for the mean and for unconditional quantiles. Moreover, our measure of (over-)confidence differs from the measure used by Reuben et al. (2017), as our tool is directly related to wage expectations whereas Reuben et al. (2017) measure overconfidence by comparing students' subjective probability of being ranked first in a mathematical task with their true probability of ranking first. Whereas Reuben et al. (2017) work with students' expected salaries across all majors, we restrict our analysis to students' wage expectations within their chosen college major, which are presumably more reliable, precise, and informative.

The remainder of the paper is organized as follows. Section 2 describes the data, shows descriptive statistics, and explains how we measure confidence. In Section 3, we describe the methods applied to decompose the gender gap in expected salaries. Section 4 presents the results of our estimations including detailed decompositions of the gender gap at the mean as well as at the deciles. Section 5 concludes.

2 Data and Descriptive Evidence

2.1 Data Set

We use data from a survey of prospective students at Saarland University, Germany, who applied for enrollment in the academic years 2011 and 2012 (henceforth Student

Survey). Prospective students submitting a complete application received an e-mail with the URL of the survey and were asked to fill in the questionnaire truthfully. Whereas in 2011 only applicants for a program in Business Studies, Humanities, Law Studies as well as Mathematics and Computer Science received the questionnaire, resulting in 500 completed surveys, in 2012 also applicants for a program in Education and Medicine were surveyed leading to 1,561 completed surveys. The sample consists of 2,061 students in total, which is a relatively large sample size compared with other studies examining students' wage expectations (e.g. Reuben et al., 2017; Webbink and Hartog, 2004).[1] The data were first used by Klößner and Pfeifer (2019), who also validated the quality of the data.[2] As Klößner and Pfeifer (2019) document, Saarland University and its students are well representative of the average university and the average student body in Germany. Figures like student/teacher ratio, gender 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 also appears in the middle of international rankings across items such as teaching/learning, research, knowledge transfer, international orientation, and regional engagement.

The prospective students first had to state the field of study and the degree (Bachelor, Master, or State Examination) they applied for. They were also asked whether they intended to do a further degree afterwards (Master, Second State Examination, or Doctoral Degree) and with which degree they intended to earn their first salary. In the next part of the survey, the prospective students were asked to answer several questions regarding monthly gross salaries.[3] First, they had to provide an estimate of the average starting salary of other persons graduating in the same field of study and with the same degree. Second, they had to state their own expected starting salary, again referring to

[1] We drop 13 prospective students with incomplete information on their wage expectations.

[2] Klößner and Pfeifer (2019) evaluate the students' knowledge of the German income tax system and focus on the difference between expected gross and net salaries. They document that prospective students tend to underestimate the progressiveness of the tax system and propose a correction for expected net salaries. Throughout our study, we use expected gross salaries.

[3] Students were provided with detailed explanations about the difference between gross and net salaries, as can be seen in an extract of the questionnaire shown in Figure 8 in the Appendix.

their chosen field and intended degree. Figure 8 in the Appendix shows an extract of the questionnaire containing these two core questions.

Furthermore, the students had to provide information about personal characteristics and family background. Specifically, we collected information on their gender; age; grade point average in secondary school; the type of secondary school they had graduated from; the federal state in which they had graduated from secondary school; whether their mother and father had a college degree and if so, their major discipline; whether they intended to live at their parents' home while studying; whether they expected to receive the public financial aid "BAfoeG"[4] and, if so, how much.

In the last part of the survey, students had to state the branch of business they intended to work in and their work experience in this branch. They also had to answer two questions regarding the value they placed on their future income: the importance of expected income for their choice of college major and the importance of receiving 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.[5] The last column of Table 1 also shows t-values for the test of equality of means between males and females.

Male students expect on average higher starting salaries both for themselves and for an average other person in the same field and with the same degree compared with female students. More specifically, males expect on average to earn 3,579 € a month at labor market entry while the corresponding number for females is 3,015 €. At the same time, the standard deviation of the expected starting salary of men is with 1,908 € as opposed to 1,558 € larger than that of women. With regard to the average earnings of others in the same field of study and with the same degree, the male estimate is 3,571 € and

[4]The Bundesausbildungsförderungsgesetz (BAfoeG, Federal Training Assistance Act) regulates student grants and loans in Germany that are granted to students with a weaker financial background.

[5]Figure 9 in the Appendix shows the densities of wage expectations.

Table 1: Descriptive Statistics

	Males	Females	t-Statistic
(a) Expected Gross Monthly Salaries (Euro)			
Own Starting Salary	3,579.12 (1,908.16)	3,014.89 (1,558.45)	7.14
Average Starting Salary	3,571.00 (1,839.61)	3,178.37 (1,620.37)	5.02
(b) Personal Characteristics			
Age (Years)	21.22 (3.85)	20.78 (3.50)	2.61
GPA Secondary School (Scale from 1 to 4)	2.31 (0.60)	2.16 (0.59)	4.91
Relevant Work Experience (Months)	4.36 (17.30)	4.23 (16.22)	0.17
(c) Field of Study (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
(d) Importance of Expected Income for Choice of Major (Percent)			
Very Strong	2.18 (14.62)	1.78 (13.24)	0.64
Strong	16.09 (0.37)	14.43 (35.16)	1.03
Neutral	37.93 (48.55)	37.01 (48.30)	0.42
Low	17.13 (37.70)	18.34 (38.71)	-0.71
Very Low	26.67 (44.25)	28.44 (45.13)	-0.89
(e) Importance of an Above-Average Salary (Percent)			
Very Important	8.97 (28.59)	5.01 (21.82)	3.41
Important	42.30 (49.43)	35.14 (47.76)	3.28
Neutral	35.29 (47.81)	42.44 (49.45)	-3.00
Unimportant	8.51 (27.91)	11.88 (32.37)	-2.53
Very Unimportant	4.94 (21.69)	5.52 (22.84)	-0.58
Observations	870	1,178	

Notes: The first two columns show means and standard deviations (in parentheses). The last column contains t-values of a test of equality of means between males and females. *Source:* Student Survey, own calculations.

significantly higher than the female estimate at 3,178 €. Interestingly, there remains a highly significant gender gap in expected earnings for others even after controlling for field of study and degree as can be seen in Table 6 in the Appendix. Hence, male and female students seem to differ in their perception of average starting salaries. There is also a gender difference in the size of the expected own salary relative to expected average salary. Whereas the expected own salary is on average higher than the expected average salary for male students, the opposite is the case for female students.

Turning to the personal characteristics of the respondents, 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 old. Both men and women, have somewhat more than four months of work experience in the branch of business they intend to work in after graduation. Male students perform worse in secondary school than their female counterparts. The final grade point average in secondary school of males is on average 2.31, whereas that of females is 2.16.[6]

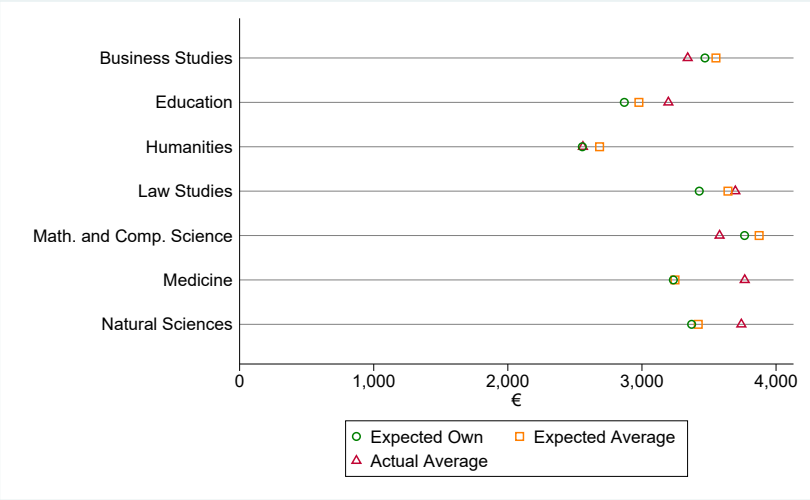
Moreover, males and females differ substantially with regard to the choice of field of study, as is also found by previous studies (e.g. Zafar, 2013; Osikominu et al., 2020). Men in the sample are significantly more likely to apply for a program in Business Studies or in Mathematics and Computer Science than women. While 16 percent of the male applicants choose a program in Mathematics and Computer Science only four percent of females do so. Contrarily, significantly larger shares of women apply for a program in Education, Humanities, or Medicine. For example, whereas 29 percent of females apply for a program in Medicine, only 19 percent of males do so.

There are also differences between male and female students in terms of the value they place on their income after graduation. Male students value an above-average salary more than female students: whereas 51 percent of males consider an above-average salary as important or very important only 40 percent of females do so. Similarly, male students are two percentage points more likely to state that future income prospects were important or very important for their choice of college major (18% of males compared with 16% of females). This suggests that men are more strongly driven by pecuniary incentives than women who seem to place a higher value on non-pecuniary aspects of college majors and employment opportunities. Similar evidence has also been documented, amongst others, in Wiswall and Zafar (2018); Osikominu et al. (2020).

Figure 1 displays mean values of expected own starting salaries and expected average starting salaries based on the Student Survey as well as actual average starting salaries

[6]In Germany, grades are scaled from 1 to 6, with 1 denoting the best grade and 4 the lowest passing grade.

Figure 1: Expected and Actual Starting Salaries



Notes: Green circles show mean values of expected own starting salaries. Orange squares show mean values of expected average starting salaries of students in the same field of study. Red triangles show actual average starting salaries. *Source:* PersonalMarkt Services GmbH (date) and Student Survey, own calculations.

differentiated by field of study.[7] Wage expectations differ considerably across these majors. While Mathematics and Computer Science students expect the highest starting salaries for themselves and for average others, Education and Humanities students expect the least. The average expected monthly starting salary is more than 1,000 € smaller for prospective students in Humanities compared with Mathematics and Computer Science. The comparison of expected starting salaries with actual starting salaries reveals that prospective students make fairly small estimation errors on average.[8] Prospective students in Business Studies as well as Mathematics and Computer Science slightly overestimate starting salaries. Prospective students in Education, Law Studies, Medicine, and Natural Sciences tend to underestimate them. Remarkably, the average expected own starting salary of Humanities students is almost identical to the actual average. Overall, it seems that prospective students possess already a sound understanding of average starting salaries in their chosen field.[9]

[7]Data on actual starting salaries are from PersonalMarkt Services GmbH (date) that offers the largest database of actual salaries for Germany.

[8]Klößner and Pfeifer (2019), who use the same data set, analyze estimation errors of students in more detail and find an average estimation error of six percent.

[9]Figure 10 in the Appendix provides further evidence on the share of men and women whose expected

2.3 Measuring (Over-)Confidence

Our data allow us to compare the expectations of students about their own starting salary, conditional on field of study and intended degree, with their expectations about the average starting salary of other students in the same field of study and with identical intended degree. Specifically, in the survey, we first elicit a respondent's expectation about the average starting salary of graduates from the same field of study and with the same degree before we elicit the starting salary that the respondent expects for him-/herself (see Figure 8 in the Appendix). In this way, we explicitly allow the respondent to anchor the estimate of their own starting salary to their perception of the labor market conditions facing an average graduate with the same degree and field of study. Therefore, we interpret individual-level differences between the two expected salaries as an indicator of a student's individual confidence and we use this information to construct various indicators of individual confidence.

In our benchmark specification, we measure a student's confidence as follows:

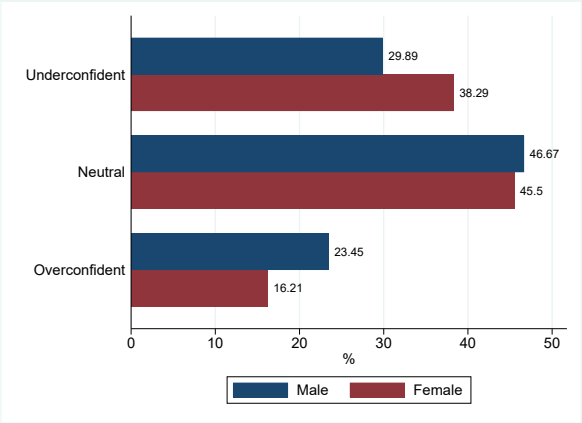
$$confidence_i = \frac{expected\ salary_{own,i}}{expected\ salary_{others,i}}\,, \tag{1}$$

which is simply the ratio of the expected own starting salary of individual i, $expected\ salary_{own,i}$, to what individual i expected as the average starting salary of all other fellow students $expected\ salary_{others,i}$. Thus, the index takes on a value larger than one if the student expects to earn more than average whereas a value below one indicates that a student expects to earn below average. As we will further substantiate in the empirical analysis we interpret a value above one as overconfidence and analogously a value below one as underconfidence.

An obvious concern with this definition of overconfidence and underconfidence may be it that it reflects differences in abilities rather than confidence per se. In fact, somebody who expects to earn more than the average other graduate in the same program may do

own salary exceeds the actual average salary.

Figure 2: Incidence of Over- and Underconfidence by Gender



Notes: The figure shows the shares of students who are overconfident ($confidence > 1$), neutral ($confidence = 1$), and underconfident ($confidence < 1$) with respect to their starting salary. *Source:* Student Survey, own calculations.

so because he/she has had above average grades in secondary school. If all students were able to realistically asses their own abilities and based their wage expectations on them, the confidence index would just measure the abilities of students. To address this concern we propose alternative measures of over- and underconfidence that take a prospective student's grade point average in secondary school into account. The GPA in secondary constitutes a salient summary measure of a student's abilities because its value and the corresponding approximate rank are well known by the students themselves and it is also an important criterion for admission to the different university programs. To be precise, we define a prospective student as overconfident who expects to earn more than the average graduate in the same program but whose GPA in secondary school is below the median in their chosen program. Analogously, we define a student as underconfident who expects to earn less than average but has a GPA in secondary school above the median in their chosen program.

Figure 2 displays the shares of male and female students who expect to earn less than average (underconfident), exactly the average (neutral), or more than the average salary (overconfident) according to our benchmark confidence index in eq. (1). The largest share of males (47 percent) and females (46 percent) expects to earn the same starting salary as their fellow students earn on average. So, almost half of all students do not expect to

Table 2: Means of Alternative Confidence Measures by Gender

	Overall	Males	Females	t-Statistic
(a) Confidence Index	1.016	1.061	0.983	3.353***
	(0.477)	(0.624)	(0.325)	
(b) Overconfident GPA adj.	0.092	0.136	0.060	5.218***
	(0.289)	(0.343)	(0.237)	
(c) Underconfident GPA adj.	0.157	0.120	0.184	-3.446***
	(0.364)	(0.325)	(0.387)	
(d) Residualized Confidence Index	0.000	0.047	-0.0347	3.602***
	(0.468)	(0.609)	(0.323)	

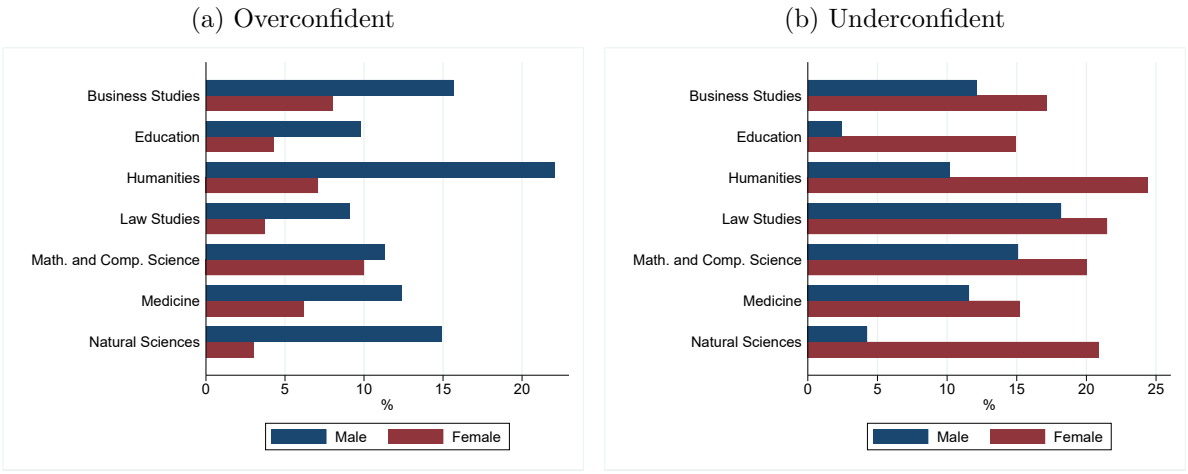
Notes: Table shows means and standard deviations (in parentheses) of alternative confidence measures. The last column shows t-values of a test of equality of means between males and females. *, ** and *** indicate statistical significance at the 10%-, 5%-, and 1%-level, respectively. *Source:* Student Survey, own calculations.

earn more or less than the average starting salary of students in their chosen program. In contrast, 38 percent of females and 30 percent of males think that they will earn less than the average salary at labor market entry while 23 percent of males and 16 percent of females expect to earn more than average. This indicates that males are more likely to be overconfident with respect to their starting salary than females, while females are more likely to be underconfident.

Row (a) of Table 2 shows the means of the confidence index in the full sample as well as by gender. While, overall, the mean is close to one, suggesting that students on average expect to earn the average starting salary of graduates from their program, men are more confident than women. Specifically, male students expect on average to earn six percent more than the expected average starting salary. Contrarily, females expect to earn on average two percent less than the expected average starting salary. This gender difference is statistically significant as the test of equality of means shows in the last column.

Rows (b) and (c) of Table 2 display the means of our alternative over- and under-confidence indicators that are defined in relation to the GPA at secondary school. They address the potential concern regarding our benchmark confidence index that students who expect to earn more than average might simply have above average ability. Even though female students performed on average better in secondary school (see Table 1),

Figure 3: GPA Adjusted Confidence by Field of Study and Gender

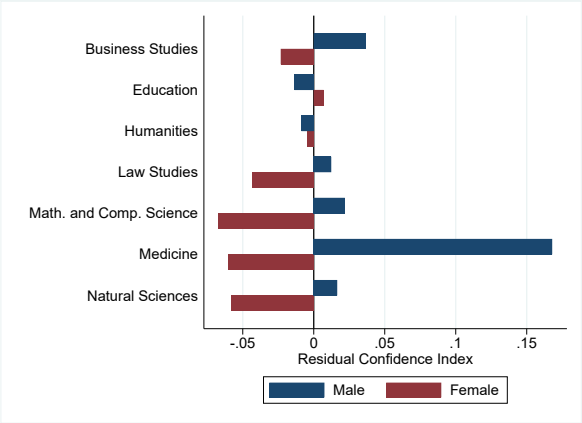


Notes: Panel (a) shows the share of students who expect to earn more than average and have a final grade in secondary school that is worse than the median grade in the chosen field of study. Panel (b) shows the share of students who expect to earn less than average and have a final grade in secondary school that is better than the median grade in the chosen field of study. *Source:* Student Survey, own calculations.

they are with 18 percent as opposed to 12 percent significantly more likely to be underconfident than male students. Moreover, the share of overconfident male students is with 14 percent more than twice as large than that of female students at six percent. Similar patterns emerge within all fields of study, as can be seen from Figure 3.

Alongside ability, there could be other plausible reasons why a student expects to earn more or less than other graduates from the same program earn on average. Prospective students may value financial aspects associated with their study choice differently or desire to work in different industries with differing pay systems. For instance, a prospective student with particularly strong preferences for non-pecuniary job amenities (e.g. flexible working hours) could rationally expect to receive a below average salary. The evidence presented in Table 1 suggests in fact that there exist important gender differences in this respect. Therefore, gender differences in the unconditional confidence index in eq. (1) and to a weaker extent also in the GPA adjusted over-/underconfidence indicators may be confounded by such additional influences of individual wage expectations. To confirm that such concerns are empirically unfounded we compute in addition a residualized confidence index that is based on the residuals from a regression of the benchmark confidence index

Figure 4: Residualized Confidence Index by Field of Study and Gender



Notes: The figure shows means of the residualized confidence index by field of study and gender. *Source:* Student Survey, own calculations.

in eq. (1) on all the control variables that we also use in our further regression analysis and that might explain rational deviations from an average expected starting salary. More precisely, we control for GPA in secondary school, parental academic background, importance of income, desired branch, age, work experience, federal state, intention to live at parents' home while studying, and whether a student expects to receive the public financial study aid BAFöG.

Row (d) of Table 2 shows the mean values of the residualized confidence index by gender. Even conditional on a broad range of determinants of expected wages men are overall significantly more confident than women, which confirms our previous unconditional results. Also within field of study, men tend to be more confident than women, as is shown in Figure 4. In Education and Humanities, however, male students are less confident than their female counterparts once we condition on other determinants of wage expectations. Moreover, the comparison across fields of study shows that male prospective students in Medicine are by far the most overconfident subgroup.[10] Female prospective students in Mathematics and Computer Science, Medicine, as well as in Natural Sciences are the least confident.

Even though we account for many observed characteristics, there might still be con-

[10]Note, however, that all of our subsequent results are robust to an exclusion of prospective medicine students.

cerns that gender differences in unobserved determinants of wage expectations affect our confidence measures. Given that females tend to work fewer hours during their lifetime, one might expect that females intend to work fewer hours after graduation and, consequently, expect lower starting salaries. The previous literature, however, suggests otherwise. Neither is there a substantial gender difference in working hours of recent university graduates (Francesconi and Parey, 2018), nor are female students less likely to expect to be in full-time employment directly after graduation (Fernandes et al., 2020). Hence, we do not believe that unobserved gender differences in intended working hours affect our confidence measure. Unobserved non-cognitive skills may be another concern. In fact, the literature on the gender gap in actual pay documents that gender differences in non-cognitive skills moderately contribute to the gender pay gap (e.g. Blau and Kahn, 2017). However, previous research on wage expectations shows that gender differences in personality traits and risk preferences do not contribute to the gender gap in expected starting salaries (Kiessling et al., 2019; Reuben et al., 2017). Therefore, we do not think that such factors have a major impact on the gender differences in our confidence measures.

Overall, the different confidence measures we use provide consistent evidence supporting our interpretation of the benchmark confidence index in eq. (1) as a measure of overconfidence and pointing towards a higher incidence of overconfidence among men than women. Our descriptive findings are thus in line with most of the previous literature on (over-)confidence (e.g. Barber and Odean, 2001; Bengtsson et al., 2005; Niederle and Vesterlund, 2007; Dahlbom et al., 2011).

3 Methodology

3.1 Oaxaca-Blinder Decomposition

To analyze the difference between the means of the own expected starting salaries of men and women we apply the well known decomposition method proposed by Oaxaca (1973)

and Blinder (1973). We use the method to compute both the aggregate decomposition of the gender gap in wage expectations into an explained and an unexplained part as well as the detailed decomposition that allows us to assess in particular the contribution of the confidence measures to the gap.

The standard Oaxaca-Blinder decomposition adopts a linear specification for the conditional expectation functions of the group-specific outcome variables. This gives rise to a linear model for the outcome variable in each group that is separable into the influence of observed and unobserved characteristics (Firpo et al., 2011). Specifically, we model the logarithm of the expected own starting salary, y_{gi}, of individual i in group $g = 0, 1$ as:

$$y_{gi} = \mathbf{x}_i \boldsymbol{\beta}_g + v_{gi} \qquad \text{for} \quad g = 0, 1\,, \tag{2}$$

where $\mathbf{x}_i$ denotes the vector of explanatory variables, including also the confidence measures and a constant, $\boldsymbol{\beta}_g$ the corresponding coefficient vector, and v_{gi} the error term with $\mathbb{E}[v_{gi} \,|\, \mathbf{x}_i] = 0$.

Define G a dummy variable equal to one if group membership is equal to one and $y \equiv (1 - G)\, y_0 + G\, y_1$. Then the overall mean difference between females ($G = 1$) and males ($G = 0$), Δ_O^μ, can be decomposed as follows (Firpo et al., 2011):[11]

$$\begin{aligned} \Delta_O^\mu &= \mathbb{E}[y \,|\, G = 1] - \mathbb{E}[y \,|\, G = 0] \\ &= \underbrace{\mathbb{E}[\mathbf{x} \,|\, G = 1] \cdot (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)}_{\Delta_U^\mu} + \underbrace{(\mathbb{E}[\mathbf{x} \,|\, G = 1] - \mathbb{E}[\mathbf{x} \,|\, G = 0]) \cdot \boldsymbol{\beta}_0}_{\Delta_E^\mu}\,. \end{aligned} \tag{3}$$

Hence, the overall mean difference between female and male wage expectations is decomposed into two components, Δ_U^μ and Δ_E^μ. The first component, Δ_U^μ, represents the unexplained part of the mean difference that captures the difference between the male and female coefficient vector $\boldsymbol{\beta}_g$. This part of the gender gap is caused by differences in the way females and males with the same characteristics, $\mathbf{x}$, form their wage expectations.

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

The second component, Δ_E^μ, represents the explained part, which captures the part of the gender gap that is due to differences in female and male characteristics. To estimate the two components Δ_U^μ and Δ_E^μ we replace $\mathbb{E}[\mathbf{x} \,|\, G = g]$ by the respective sample mean, $\bar{\mathbf{x}}_g$, and plug in an estimate of the respective coefficient vector $\hat{\boldsymbol{\beta}}_g$, which we obtain from OLS regressions of eq. (2) for females and males (Firpo et al., 2011). Specifically, we compute an estimate of the overall mean difference, $\hat{\Delta}_O^\mu$, as follows:

$$\hat{\Delta}_O^\mu = \bar{y}_1 - \bar{y}_0 = \underbrace{\bar{\mathbf{x}}_1(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_0)}_{\hat{\Delta}_U^\mu} + \underbrace{(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_0)\hat{\boldsymbol{\beta}}_0}_{\hat{\Delta}_E^\mu}. \tag{4}$$

Because of the additive linear specification for expected starting salaries (eq. 2), components $\hat{\Delta}_U^\mu$ and $\hat{\Delta}_E^\mu$ in eq. (4) are additive in the contributions of the individual explanatory variables, which we exploit to obtain the detailed decomposition, see Firpo et al. (2011) for further details.

3.2 Decomposing Unconditional Quantiles

To gain a more detailed understanding of the gender gap not only at the mean but also at other parts of the unconditional distribution of own expected starting salaries we implement in addition the approach proposed by Firpo et al. (2009, 2018). The basic idea of this method is to model the recentered influence function (RIF) of the τ-th quantile, Q_τ, of the outcome variable y_g, $g = 0, 1$, which can be expressed as

$$\text{RIF}(y_g; Q_\tau) = Q_\tau + \text{IF}(y_g; Q_\tau) = Q_\tau + \frac{\tau - \mathbb{1}\{y_g \leq Q_\tau\}}{f_{y_g}(Q_\tau)}, \tag{5}$$

where $\text{IF}(y_g; Q_\tau)$ denotes the influence function of quantile Q_τ, $f_{y_g}(\cdot)$ the marginal density of y_g, and $\mathbb{1}\{\cdot\}$ an indicator function taking on the value one if the condition in the argument is true (Firpo et al., 2009). As, by definition, the expected value of the influence function is equal to zero, i.e. $\mathbb{E}[\text{IF}(y_g; Q_\tau)] = \int \text{IF}(y_g; Q_\tau) \cdot dF(y_g) = 0$, the expected value of the RIF is equal to Q_τ itself (Firpo et al., 2018). Using the law of iterated expectations,

the quantile Q_τ can therefore be written as

$$Q_\tau = \mathbb{E}[\mathbb{E}(\text{RIF}(y_g; Q_\tau) \,|\, \mathbf{x})] = \int \mathbb{E}[\text{RIF}(y_g; Q_\tau) \,|\, \mathbf{x}] \cdot dF_{\mathbf{x},g}(\mathbf{x}) \,,$$

where $F_{\mathbf{x},g}$ denotes the marginal distribution of $\mathbf{x}$ in group g.

We adopt a linear specification for the conditional expectation of the RIF as a function of the explanatory variables $\mathbf{x}$:

$$\mathbb{E}[\text{RIF}(y_{gi}; Q_\tau) \,|\, \mathbf{x}_i] = \mathbf{x}_i \boldsymbol{\gamma}_{\tau,g} \,. \tag{6}$$

This implies that we obtain a model for $\text{RIF}(y_{gi}; Q_\tau)$ that is additively separable in the influence of observables and unobservables:

$$\text{RIF}(y_{gi}; Q_\tau) = \mathbf{x}_i \boldsymbol{\gamma}_{\tau,g} + v_{\tau gi} \,, \tag{7}$$

with $\mathbb{E}[v_{\tau gi} \,|\, \mathbf{x}_i] = 0$, which can be estimated using an OLS regression. Further, we obtain the unconditional expectation of $\text{RIF}(y_g; Q_\tau)$, which is the τ-th quantile of y_g, if we evaluate eq. (6) at the expected value of $\mathbf{x}$ in group $G = g$, $\mathbb{E}[\mathbf{x} \,|\, G = g]$. Consequently, the j-th coefficient, $\gamma_{\tau,g,j}$, in eq. (7) can be interpreted as the effect of a unit increase of the mean of x_j in group g on the τ-th unconditional quantile of the outcome variable in group g. Hence, the gender gap at a specific quantile of the unconditional distribution of expected starting salaries can be decomposed as in the standard Oaxaca-Blinder decomposition when the dependent variable in the Oaxaca-Blinder decomposition, y_g, is replaced by the RIF at the respective quantile, $\text{RIF}(y_g; Q_\tau)$. To implement the Oaxaca-Blinder decomposition of the RIF, we obtain estimates of $\text{RIF}(y_g; Q_\tau)$ in eq. (5) for each group by inserting the respective sample quantile, $\hat{Q}_\tau$, and an estimate of the density at that quantile, $\hat{f}_{y_g}(\hat{Q}_\tau)$ (Firpo et al., 2009).[12]

[12] We use a kernel density estimator with a Gaussian kernel and a bandwidth of 0.1 for this purpose.

3.3 Identification

The goal of this paper is to identify and estimate the explained and unexplained part of the overall gender gap as well as the specific contribution of the confidence measures to these two components of the total gap. Given the linear specifications of the conditional expectation functions of the log of expected starting salaries[13] in the Oaxaca-Blinder decomposition the key identifying assumption is that conditional on the explanatory variables the group-specific outcome variable, y_g, is mean independent of group membership, i.e. equivalently

$$\mathbb{E}[v_g \,|\, \mathbf{x}, G] = \mathbb{E}[v_g \,|\, \mathbf{x}] = 0\,, \quad g = 0, 1\,. \tag{8}$$

This assumption parallels the so-called conditional independence or ignorability assumption in the treatment effects literature (Firpo et al., 2011).

Similarly, we assume in the RIF regressions that locally, at the τ-th Quantile of y_g,

$$\mathbb{E}[v_{\tau g} \,|\, \mathbf{x}, G] = \mathbb{E}[v_{\tau g} \,|\, \mathbf{x}] = 0\,, \quad \forall \tau \in [0, 1]\,,\ g = 0, 1\,. \tag{9}$$

Thus, to be able to interpret the explained and unexplained part of the decomposition truly as composition and wage structure effect, we need to condition on all variables that are correlated with both the group-specific outcome variables and group membership. In our application, we include a broad range of explanatory variables that we have selected based on economic reasoning and prior evidence on the determinants of expected starting salaries among prospective university students. As outlined in Section 2, we consider variables intended to capture heterogeneity in the prospective students' tastes for academic disciplines, academic abilities, attitudes towards monetary returns from studying, economic and academic background of parents, preferences for desired jobs as well as knowledge about the desired branch of business. Importantly, even though we do not

[13] In a sensitivity analysis, we verify that the linear specifications for the conditional expectation functions of the log of expected starting salaries as well as the RIF of the log of expected starting salaries are justified empirically using the reweighted regression approach suggested in Firpo et al. (2011) (see Appendix C).

have access to detailed controls for the underlying aspects of personality and preferences like competitiveness or personal style in pay negotiations, to satisfy assumption (8) and (9), respectively, it suffices to condition on variables that can proxy the underlying facets well enough to achieve conditional independence of group-specific outcomes and group membership. As an example, while we cannot directly control for competitiveness and personal style in pay negotiations, we know that students who choose a program in Education and indicate that pay is not important to them have similar competitiveness and negotiation styles and are typically less competitive and less bold in pay negotiations than students majoring in Business who state that pay is important to them.[14]

To identify and estimate the contribution of (over-)confidence to the gender gap in wage expectations, we rely on exogeneity of the explanatory variables with respect to the group specific error terms as stated in the second part of eq. (8) and (9). These exogeneity conditions require us to control for all determinants of expected starting salaries of men and women that are also correlated with (over-)confidence. By the arguments provided in this and in Section 2.3, we are confident that our data and model specification are rich enough to capture the potential confounders. Moreover, our benchmark confidence index is computed as the ratio of the starting salary expected for oneself to that expected for others. Hence, proportional misperceptions of the employment prospects after graduation that equally affect the expectation for oneself and for others cancel out.

To assess the robustness of our results on the role of (over-)confidence to selection on unobservables we implement a method recently proposed by Kiviet (2016, 2020) that allows us to infer the degree of endogeneity of the benchmark confidence index that would render its effect on the log expected salary statistically insignificant.

[14]See also Dale and Krueger (2002) who apply a similar strategy to estimate the causal return to attending a more selective college in the US.

Table 3: OLS Regressions of Log Expected Starting Salary

	Pooled		Males		Females	
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.1717***	-0.0975***				
	(0.0247)	(0.0251)				
Confidence Index		0.2020***	0.1447***	0.1697***	0.3082***	0.2902***
		(0.0249)	(0.0296)	(0.0293)	(0.0488)	(0.0479)
Controls		✓		✓		✓
R-squared	0.0231	0.1591	0.0268	0.1877	0.0327	0.1547
Observations	2,048	2,048	870	870	1,178	1,178

Notes: The dependent variable is the logarithm of the expected gross starting salary. In Columns (2), (4), and (6), we control in addition for the degree applied for, the field of study, the degree with which the student intends to earn the first salary, importance of income, human capital variables, personal as well as family background, and the survey year. Standard errors are shown in parentheses. *, ** and *** indicate statistical significance at the 10%-, 5%-, and 1%-level, respectively. *Source:* Student Survey, own calculations.

4 Estimation Results

4.1 The Gender Gap in Wage Expectations and the Effect of Confidence

Columns (1) and (2) of Table 3 show the gender gap in expected starting salaries obtained from regressions of the log expected starting salary on a female dummy once without and once with further controls. The estimates suggest that female students expect to earn less than males upon graduation, with the difference being equal to 17.2 percent and highly significant in the specification excluding further controls. The raw gap we find is very similar in size to the gender gap in expected starting salaries of 18.1 percent documented in Kiessling et al. (2019).

After controlling for the confidence index as well as other explanatory variables the gender gap declines by 7 percentage points to 9.8 percent, see column (2) of Table 3. This remaining gender gap can 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 and confidence. The comparison of the estimates of the gender gap in columns (1) and (2) shows that more than a third of the gender gap in wage

expectations can be explained by differences in observed characteristics between male and female students. Nevertheless, there remains a sizable unexplained gender gap in wage expectations that is still highly significant.

Turning to the coefficient estimate of the confidence index in column (2), we see that it has a positive and significant impact on wage expectations. A ten percentage point increase in confidence increases the expected starting salary by two percent on average, holding all other factors constant. Thus, a higher level of confidence translates into higher wage expectations, a finding in line with the previous literature (Reuben et al., 2017, e.g.).

In the remaining columns of Table 3 display the estimates obtained from separate regressions in the male and female subsamples. While in Columns (3) and (5) the dependent variable is regressed on a constant and the confidence index only, Columns (4) and (5) include all additional control variables. Our Blinder-Oaxaca decomposition of the mean uses the estimates in column (4) to form the counterfactual wage structure. According to the estimate in column (3), a ten percentage points increase in confidence increases the estimated starting salary by 1.4 percent among males, whereas the corresponding estimate for females is with three percent more than twice as large (column (5)). When the additional control variables are included in columns (4) and (6), respectively, the coefficient estimates of the confidence index, change slightly and the gender difference in the effect of confidence on starting salaries becomes somewhat smaller. Taken together the evidence in Table 3 and Table 2 suggest that male and female students differ not only in the average level of confidence, but confidence also affects their wage expectations differently. Thus, we expect that confidence contributes both to the composition and the wage structure effect of the gender gap in wage expectations.

4.2 Decomposition of the Gender Gap at the Mean

4.2.1 Benchmark Specification

Table 4 shows the results of the Oaxaca-Blinder decomposition of expected starting salaries at the mean. We use the male coefficient vector to form the counterfactual distri-

Table 4: Oaxaca-Blinder Decomposition at the Mean

	(1) Decomposition Starting Salary	(2) Relative Impact (in %)
(a) Aggregate Decomposition		
Mean of Males	8.0467***	
	(0.0191)	
Mean of Females	7.8750***	
	(0.0164)	
Gender Gap	0.1717***	
	(0.0252)	
Explained Part	0.0862***	50.17
	(0.0175)	
Unexplained Part	0.0856***	49.83
	(0.0276)	
(b) Contributions of Covariates to the Explained Part		
Confidence Index	0.0132***	7.68
	(0.0045)	
Field of Study	0.0511***	29.73
	(0.0151)	
Income Importance	0.0142**	8.28
	(0.0059)	
Other	0.0077	4.48
	(0.0133)	
(c) Contributions of Covariates to the Unexplained Part		
Confidence Index	-0.1185**	-68.99
	(0.0552)	
Field of Study	-0.0072	-4.16
	(0.0683)	
Income Importance	0.0483	28.14
	(0.0440)	
Other	-0.1098	-63.92
	(0.2444)	
Constant	0.2727	158.77
	(0.2575)	
Observations	2,048	

Notes: The dependent variable is the logarithm of the expected gross starting salary. The male coefficient vector is used as the reference. Standard errors are calculated as described by Jann (2008) and are shown in parentheses. *, ** and *** indicate statistical significance at the 10%-, 5%-, and 1%-level, respectively. *Source*: Student Survey, own calculations.

bution.[15] The estimates in the panel (a) reveal that 8.6 percentage points of the overall gender gap of 17.2 percent can be attributed to differences in observed characteristics between males and females. Another 8.6 percentage points of the gap remain unexplained and are attributable to differences in the structure of male and female wage expectations. This means that 50 percent of the gender gap in expected own starting salaries can be explained by differences in the composition of male and female students.

The results of the detailed decomposition in panel (b) show that the difference in confidence between male and female students contributes significantly 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 confidence. This means in relative terms that about 8 percent of the overall gender gap (15 percent of the explained gap) in the expected starting salary is due to the higher confidence of male students. This finding is in line with Reuben et al. (2017) who also document that gender differences in overconfidence are partly responsible for higher expected salaries of male students.[16] Gender differences in the choice of field of study together contribute 5.1 percentage points to the gender gap, which corresponds to roughly 30 percent of the overall gap or 60 percent of the explained gap. Thus, around a third of the gap can be explained by differential sorting of males and females into 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 Mathematics and Computer Science, which are the fields with the highest expected starting salaries. Differential sorting into university majors, occupations and industries has been documented as an important explanation for the gender gap in both expected as well as realized earnings also in the existing literature (Collischon, 2019, e.g.). Comparing the magnitude of the effect of confidence with that of field of study, we see that the former contributes about a quarter of the contribution of the field of study to

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

[16]Specifically, Reuben et al. (2017) find that gender differences in overconfidence and competitiveness together explain 18 percent of the gender gap in earnings expectations.

explaining the overall gender gap. This suggests that confidence plays a substantive role for the gender gap in wage expectations.

The third characteristic of students that significantly contributes to the explained part of the gender gap in expected starting salaries are students' views on the importance of income. We capture income importance with two sets of dummy variables, one on the importance of receiving an above-average salary and another one on the importance of pecuniary returns for the choice of major. 1.4 percentage points of the gap can be attributed to gender differences in the set of variables measuring 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 (Zafar, 2013, e.g.). 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 explanatory variables in panel (c) of Table 4 shows how much of the gender gap is due to gender differences in the effect that a characteristic has on the structure of wage expectations. Of all the explanatory variables included, only the gender difference in the coefficient of the confidence index has an effect on the unexplained gap that is statistically different from zero, but negative. This means that a unit increase in the mean level of confidence has a larger positive effect on the mean expected starting salary of women than men, holding constant all other determinants of expected starting salaries (see also columns (4) and (6) of Table 3). Finally, we document a substantial contribution of the difference in the intercept of male and female wage expectation structures to the unexplained part of the gender gap, which captures unobserved factors. Hence, our findings do not suggest that women have lower expectations regarding their starting salaries because they expect lower returns than men to the characteristics considered and especially confidence.

Table 5: Detailed Decomposition Results for Alternative Confidence Measures

	(1) Benchmark	(2) Quadratic Index	(3) Overconfident GPA Adj.	(4) Underconfident GPA Adj.	(5) Over- & Under-conf. GPA Adj.
(a) Aggregate Decomposition					
Gender Gap	0.1717***	0.1717***	0.1550***	0.1550***	0.1550***
	(0.0252)	(0.0252)	(0.0286)	(0.0286)	(0.0286)
Explained Part	0.0862***	0.0873***	0.0873***	0.0834***	0.0939***
	(0.0175)	(0.0175)	(0.0209)	(0.0209)	(0.0212)
Unexplained Part	0.0856***	0.0844***	0.0677**	0.0717**	0.0612*
	(0.0276)	(0.0275)	(0.0323)	(0.0325)	(0.0323)
(b) Contributions of Covariates to the Explained Part					
Confidence	0.0132***	0.0156***			
	(0.0045)	(0.0051)			
Overconfidence			0.0204***		0.0199***
			(0.0063)		(0.0062)
Underconfidence				0.0107**	0.0098*
				(0.0053)	(0.0051)
Field of Study	0.0511***	0.0506***	0.0567***	0.0616***	0.0604***
	(0.0151)	(0.0151)	(0.0180)	(0.0182)	(0.0180)
Income Importance	0.0142**	0.0136**	0.0086	0.0092	0.0076
	(0.0059)	(0.0058)	(0.0072)	(0.0072)	(0.0071)
Other	0.0077	0.0075	0.0016	0.0019	-0.0039
	(0.0133)	(0.0133)	(0.0166)	(0.0167)	(0.0168)
Observations	2,048	2,048	1,564	1,564	1,564

Notes: The dependent variable is the logarithm of the expected gross starting salary. In all columns the male coefficient vector is used as the reference vector. Column (1) is our benchmark specification from Table 4 that includes the confidence index linearly. Column (2) includes in addition the square of the confidence index. Column (3) replaces the confidence index by a dummy for GPA adjusted overconfidence. Column (4) replaces the confidence index by a dummy for GPA adjusted underconfidence. Column (5) includes both GPA adjusted over- and underconfidence dummies. Standard errors are calculated as described by Jann (2008) and are shown in parentheses. *, ** and *** indicate statistical significance at the 10%-, 5%-, and 1%-level, respectively. *Source:* Student Survey, own calculations.

4.2.2 Alternative Specifications and Robustness

Table 5 shows results of the Oaxaca-Blinder decomposition using different measures of confidence. For comparison, our benchmark decomposition results are shown in Column (1). The specification in column (2) includes in addition the square of the confidence index. In this way, we can exmine the robustness of our results to deviations from assuming linearity of log expected salaries in confidence. The results in column (2) suggest, that allowing for a nonlinear effect of confidence has only minor effects on the results. The contribution of confidence to the gender gap appears even a bit larger compared with our benchmark specification. The estimates presented in columns (3) to (5) are based on the GPA adjusted confidence indicators described in Section 2.3. The dummy for GPA adjusted overconfidence (underconfidence) is equal to one if somebody has a confidence index larger than one (below one), while their GPA at secondary school is below (above) the median in the chosen field of study. Using these dummies rather than the confidence index addresses the concern that the confidence index reflects differences in abilities rather than differences in confidence. It further allows us to address concerns related to the functional form of the confidence index that contains in the numerator the own expected starting salary (in Euro). As can be seen from column (3), the contribution of the GPA adjusted overconfidence indicator to the gender gap is highly significant and even larger than the contribution of the confidence index in our baseline model. Column (4), shows the decomposition results using the GPA adjusted indicator for underconfidence. Gender differences in underconfidence do also significantly contribute to the gap, but less than differences in overconfidence. Finally, column (5) shows the decomposition results when we include both indicator variables. Here, only the contribution of overconfidence remains statistically significant. These results suggest that the contribution of gender differences in the mean level of the confidence index to the gender gap is driven by overconfidence of men rather than underconfidence of women.

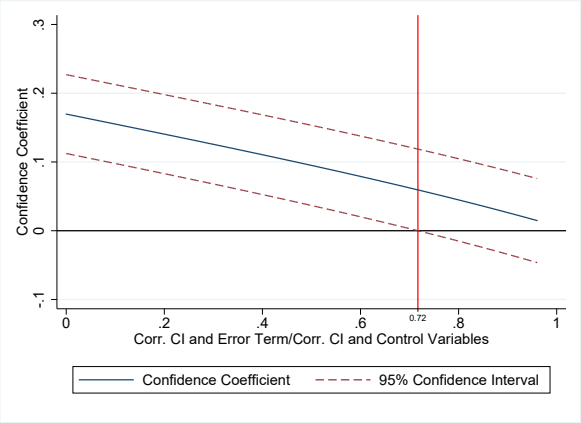
Overall, we conclude that the decomposition results are robust to the use of different confidence measures. The estimated contribution of confidence to the explained part of

the gender gap in the expected starting salary remains stable no matter which variant of the confidence measure is used. If anything, the estimated contribution gets larger when we use alternative confidence measures. Moreover, neither the size of the explained part nor the contributions of the other observed characteristics change dramatically when a different confidence measure is used. In addition, we conclude that the contribution of confidence to the gender gap mainly stems from gender differences in overconfidence.

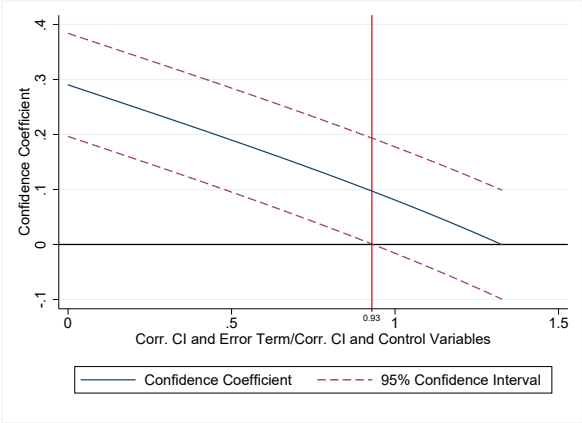
Finally, as discussed in Section 3.3 the key assumption for identification of the true contribution of confidence to the explained and unexplained gap, is that confidence is exogenous in the male and female models for the wage structure. While we are confident that our set of control variables includes all relevant aspects, we want to examine how robust our results are to a potential endogeneity of the confidence index. For this purpose, we apply the method recently proposed by Kiviet (2016, 2020) who suggests to compute point estimate and standard error of the coefficient attached to a regressor under an assumed degree of correlation between the regressor and the error term. This yields a coefficient estimate and confidence interval that are adjusted for the supposed endogeneity. By varying the assumed values of the correlation, one can examine until which value of the correlation between the regressor and the error term the estimated coefficient remains statistically different from zero. In our case, negative correlations between the confidence index and the error term would imply a higher adjusted point estimate than under the assumption of exogeneity (Kiviet, 2016). Thus, we vary the assumed value of the correlation between the confidence index and the error term in the male and female regression models on the positive range. To be able to evaluate the magnitude of the assumed values of the correlation between the confidence index and the error term, we express it relative to the correlation between the confidence index and the other control variables in the regression model for the expected starting salary. Thus, we regress the confidence index on all the other control variables and compute the correlation between the confidence index and the fitted values of this regression, which amounts to 0.31 in the male subsample and to 0.23 in the female subsample. Panels (a) and (b) of Figure 5

Figure 5: Sensitivity of Confidence to Correlation with the Error Term

(a) Male Model



(b) Female Model



Notes: The figure shows adjusted OLS estimates of the coefficient on the confidence index in the male and female model obtained under an assumed value of the correlation between the confidence index and the error term given on the horizontal axis (Kiviet, 2016, 2020). The horizontal axes show the assumed correlation between the confidence index and the error term relative to the correlation between the confidence index and the other control variables. The red vertical line indicates the value of the relative correlation at which the estimated coefficient is just significant at the 5% level. *Source:* Student Survey, own calculations.

depict the estimated coefficient of the confidence index with corresponding 95% confidence interval as a function of the assumed relative correlation between confidence and the error term in the male and female model, respectively. The assumed relative correlation shown on the horizontal axes is obtained as the correlation between the confidence index and the error term divided by the correlation between the confidence index and the other control variables. In Figure 5, the intersections of point estimates and confidence intervals with the vertical axis correspond to the estimates under assumed exogeneity of confidence shown in columns (4) and (5) of Table 3. According to panel (a), the coefficient of the confidence index remains statistically significant up to a relative correlation of 72 percent in the male model and up to a relative correlation of 93 percent in the female model shown in panel (b), as indicated by the red vertical lines in Figure 5. Thus, if in the female model the correlation between confidence and the error term attained a value about as large as the correlation between confidence and the other control variables, then the contribution of confidence to the female wage structure would turn statistically insignificant.

4.3 Decomposition of the Gender Gap at Unconditional Quantiles

As the gender gap in the expected starting salary and its causes may differ at different points in the distribution of expected starting salaries we apply the RIF-regression approach (Firpo et al., 2009), described in Section 3.2. The results are shown in Table 6 as well as in Figures 6 and 7. The gender gap is evaluated at every decile of the wage expectation distribution. Panel (a) of Figure 6 as well as panel (a) of Table 6 show the total gender gap in the expected starting salaries along the deciles of the male and female distributions. The gender gap is present at all deciles and varies rather unsystematically, being 13.5 percent at the first decile, 19.5 percent at the median, and 17.2 percent at the ninth decile.[17] In comparison, the mean of expected starting salaries of males is 17.2 percent higher than that of females (dashed line in panel (a)). So, whereas the gender gap at the bottom of the distribution of expected starting salaries is considerably smaller than at the mean it is larger at the center of the distribution. In sum, we do not find evidence of a glass ceiling (wider gap at the top) nor of a sticky floor (wider gap at the bottom). However, the reason for this could be that our sample only includes prospective college students, i.e. a group of highly skilled individuals.

Panel (b) of Figure 6 and panel (a) of Table 6 show how much of the gender gap at the different deciles of the unconditional distribution of expected starting salaries can be explained by gender differences in observed characteristics. In absolute terms, the explained part of the gender gap is with values between seven and eight percentage points rather constant across the middle and top deciles and with up to ten to eleven percentage points somewhat higher at the first two deciles. In comparison, the explained part of the gender gap at the mean is 8.6 percentage points. At the first decile, the explained part is largest even though the overall gender gap in the expected starting salary is smallest. Contrarily, the explained part is smallest at the seventh decile where the overall gender gap is fairly large. Hence, although the explained part is rather constant across deciles

[17] Kiessling et al. (2019) find larger gaps at the center and bottom of the distribution than at the top.

Table 6: Detailed Decomposition at the Deciles of the Log Expected Starting Salary

	(1) 10%	(2) 20%	(3) 30%	(4) 40%	(5) 50%
(a) 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	0.1138***	0.0998***	0.0806***	0.0841***	0.0748***
	(0.0421)	(0.0286)	(0.0228)	(0.0226)	(0.0194)
Unexplained Part	0.0216	0.0922*	0.0993***	0.0725**	0.1202***
	(0.0821)	(0.0475)	(0.0367)	(0.0329)	(0.0299)
(b) Contributions of Covariates to the Explained Part					
Confidence Index	0.0176***	0.0115***	0.0067*	0.0077***	0.0063**
	(0.0062)	(0.0044)	(0.0037)	(0.0029)	(0.0026)
Field of Study	0.0534	0.0609**	0.0461**	0.0561***	0.0500***
	(0.0360)	(0.0245)	(0.0191)	(0.0193)	(0.0168)
Income Importance	0.0220	0.0076	0.0074	0.0050	0.0035
	(0.0152)	(0.0101)	(0.0072)	(0.0074)	(0.0069)
Other	0.0209	0.0198	0.0205	0.0153	0.0149
	(0.0346)	(0.0222)	(0.0171)	(0.0176)	(0.0149)
(c) Contributions of Covariates to the Unexplained Part					
Confidence Index	-0.1972	-0.1692	-0.2164***	-0.2048***	-0.1322*
	(0.1688)	(0.1065)	(0.0808)	(0.0754)	(0.0712)
Field of Study	0.0737	-0.1015	-0.0406	-0.0606	-0.0333
	(0.1775)	(0.1093)	(0.0877)	(0.0820)	(0.0758)
Income Importance	-0.0282	-0.0585	0.0002	-0.0339	-0.0431
	(0.1141)	(0.0705)	(0.0608)	(0.0548)	(0.0467)
Other	0.3344	0.0052	0.1649	0.0725	0.1014
	(0.7767)	(0.3918)	(0.3300)	(0.3113)	(0.2826)
Constant	-0.1611	0.4163	0.1913	0.2992	0.2273
	(0.7966)	(0.4293)	(0.3449)	(0.3214)	(0.3012)
Observations	2,048	2,048	2,048	2,048	2,048

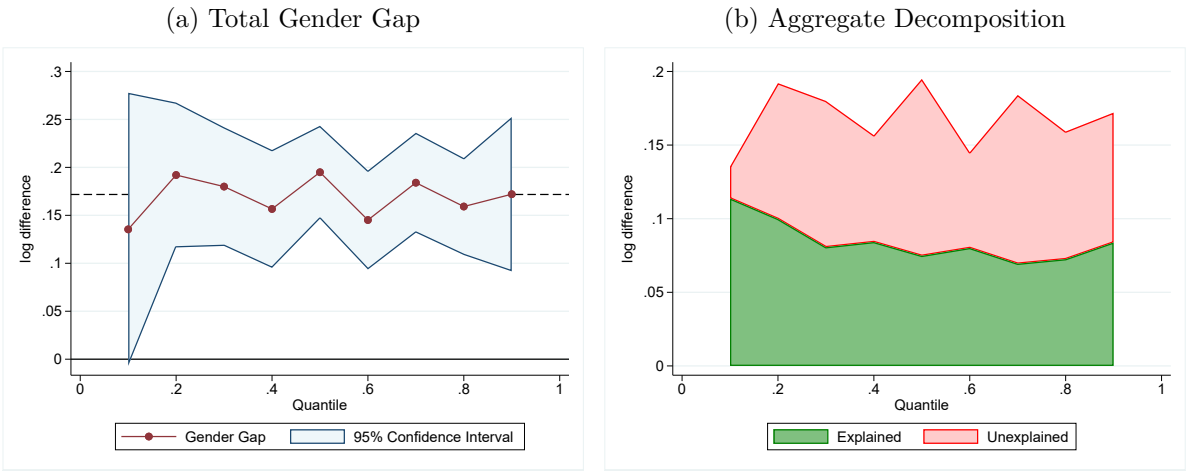
Notes: The dependent variable is the estimated RIF at the respective decile of the log expected starting salary. Males are chosen as the reference group. Bootstrap standard errors (400 replications) are shown in parentheses. *, ** and *** indicate statistical significance at the 10%-, 5%-, and 1%-level, respectively. *Source:* Student Survey, own calculations.

Table 5: Detailed Decomposition at the Deciles of the Log Expected Starting Salary *(continued)*

	(6) 60%	(7) 70%	(8) 80%	(9) 90%
(a) 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	0.0801***	0.0694***	0.0725***	0.0839***
	(0.0178)	(0.0184)	(0.0201)	(0.0259)
Unexplained Part	0.0650**	0.1146***	0.0866***	0.0882**
	(0.0305)	(0.0315)	(0.0281)	(0.0431)
(b) Contributions of Covariates to the Explained Part				
Confidence Index	0.0069***	0.0080***	0.0085**	0.0144**
	(0.0025)	(0.0027)	(0.0035)	(0.0065)
Field of Study	0.0623***	0.0601***	0.0577***	0.0390**
	(0.0153)	(0.0159)	(0.0161)	(0.0195)
Income Importance	0.0082	0.0109*	0.0136**	0.0203**
	(0.0067)	(0.0061)	(0.0068)	(0.0089)
Other	0.0027	-0.0095	-0.0072	0.0101
	(0.0139)	(0.0150)	(0.0153)	(0.0210)
(c) Contributions of Covariates to the Unexplained Part				
Confidence Index	-0.0934	-0.0979	-0.1251	-0.1016
	(0.0740)	(0.0739)	(0.0765)	(0.1154)
Field of Study	-0.0726	-0.0518	-0.0345	0.1762*
	(0.0710)	(0.0644)	(0.0748)	(0.0934)
Income Importance	0.0251	0.0475	0.0645	0.1108
	(0.0440)	(0.0458)	(0.0503)	(0.0748)
Other	0.1652	0.2213	0.3522	-0.0622
	(0.2562)	(0.2622)	(0.2943)	(0.3694)
Constant	0.0407	-0.0045	-0.1704	-0.0351
	(0.2712)	(0.2636)	(0.3027)	(0.3965)
Observations	2,048	2,048	2,048	2,048

Notes: The dependent variable is the estimated RIF at the respective decile of the log expected starting salary. Males are chosen as the reference group. Bootstrap standard errors (400 replications) are shown in parentheses. *, ** and *** indicate statistical significance at the 10%-, 5%-, and 1%-level, respectively. *Source:* Student Survey, own calculations.

Figure 6: Aggregate Decomposition at the Deciles of the Log Expected Starting Salary

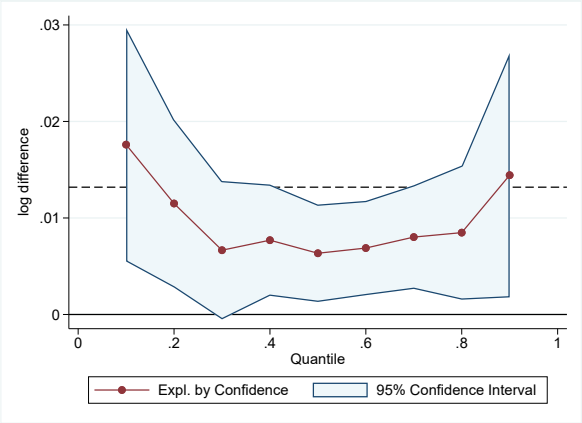


Notes: Males are chosen as the reference group. The pointwise confidence interval in panel (a) is based on bootstrap standard errors (400 replications). The dashed line in panel (a) shows the gender gap at the mean. *Source:* Student Survey, own calculations.

in absolute terms its relative importance varies substantially. Whereas 84 percent of the gender gap at the first decile can be explained by differences in observed characteristics, less than 38 percent of the gap can be explained at the seventh decile.

Figure 7 and panel (b) of Table 6 show the contribution of the confidence index to the explained part of the gender gap in wage expectations along the deciles of the unconditional distribution of expected starting salaries. The contribution of the confidence index is positive and significantly different from zero at every decile. Thus, decreasing the mean of confidence from the level of males to that of females increases the gender gap in the expected starting salary at every decile. Depending on the decile, the portion of the gap that can be attributed to gender differences in confidence lies between 1.8 and 0.6 percentage points, with higher values in the tails of the distribution than in the center. The contrast between the tails and the center of the distribution is even larger in relative terms. Whereas gender differences in confidence are responsible for 13 percent of the gap at the first decile and for eight percent of the gap at the ninth decile, they are only responsible for three percent of the gap at the median. This U-shape pattern suggests that differences in male and female confidence contribute primarily to the gender gap in the mean (as seen also in Section 4.2) rather than to the gender gap in the dispersion of

Figure 7: Contribution of the Confidence Index



Notes: The dependent variable is the estimated RIF at the respective decile of the log expected starting salary. Males are chosen as the reference group. Pointwise confidence bands are based on bootstrapped standard errors (400 replications). The dashed line shows the contribution of confidence to the explained part of the gender gap at the mean. *Source:* Student Survey, own calculations.

expected starting salaries as measured by the difference between the ninth and the first decile.

Further, the results in panel (b) of Table 6 indicate that the contribution of differential sorting into fields of study is with four to six percentage points consistently high throughout the deciles. This pattern matches the effect of field of study in the decomposition at the mean and suggests that differential sorting into fields of study does not contribute to gender differences in the dispersion of expected starting salaries.

Turning to the contribution of confidence to the unexplained gap along the deciles of expected starting salaries shown in panel (c) of Table 6, we see that the point estimates are negative across all deciles, but substantially larger in absolute value and only statistically significant in the lower half of the distribution. This suggests that, especially in the lower half of the distribution, a unit increase in the average level of confidence increases the respective decile of the female expected salary more strongly than that of men which reduces the gender gap. This finding for the unconditional deciles parallels our results on the structure effect of confidence at the mean of expected starting salaries (see Section 4.2).

5 Conclusion

Based on large and informative survey data on prospective students at Saarland University, Germany, we propose new measures to assess an individual's (over-)confidence with respect to their expected salary after graduation. We then analyze the role of (over-)confidence in shaping the gender gap in wage expectations of prospective university students.

Our results indicate that the average expected starting salary of women is 17 percent lower than that of men. At the same time and in line with previous studies, we find that male students are significantly more overconfident than female students (Barber and Odean, 2001; Bengtsson et al., 2005). According to our benchmark confidence index, the gender gap in confidence equals almost eight percent. Moreover, confidence has a positive effect on the expected starting salary. Together these results indicate already that differences in confidence contribute partly to the observed gap in male and female expected starting salaries. Going into details, the results of the Oaxaca-Blinder decomposition at the mean show that 7.7 percent of the overall gender gap (and 15 percent of the explained gap) in expected starting salaries is attributable to the fact that male students are more confident than females. At the same time, gender differences in the wage structure effect of confidence contribute negatively to the overall gender gap suggesting that a unit increase in the mean of confidence increases the mean of the female expected starting salary more strongly than the male one. Alternative decompositions that include the GPA adjusted indicators for over- and underconfidence instead of the benchmark confidence index reveal that the contribution of gender differences in confidence to the explained gap is mainly driven by gender differences in overconfidence as opposed to underconfidence. Furthermore, the decompositions at the unconditional deciles of male and female expected starting salaries show that, while the gender gap is present at all deciles, there is no evidence for sticky floor (wider gap at the bottom) nor glass ceiling effects (wider gap at the top). The contribution of the confidence index to the explained gap, in contrast, exhibits a U-shaped pattern along the deciles. This suggests that gender differences in

confidence do not contribute to the gender gap in the dispersion of expected starting salaries as measured by the difference between the ninth and first decile. Finally, our results show that differential sorting of men and women into fields of study accounts for 30 percent of the overall gap (60 percent of the explained gap) at the mean and for similarly large shares of the gaps at the deciles of expected starting salaries. This pattern parallels the findings in the literature on realized wages that generally attribute the most important share of the explained gap to differential sorting into occupations and industries.

Our findings raise the question how interindividual differences in (over-)confidence emerge and whether and how it is possible for parents and educators to support the formation of confidence in young people. If confidence is malleable, policy measures aimed at raising the confidence of girls and young women in particular might help to reduce gender differences in pay.

References

Antonczyk, D., Fitzenberger, B., and Sommerfeld, K. (2010). Rising wage inequality, the decline of collective bargaining, and the gender wage gap. *Labour Economics*, 17(5):835–847.

Arcidiacono, P., Hotz, V. J., Maurel, A., and Romano, T. (2020). Ex Ante Returns and Occupational Choice. *Journal of Political Economy*, forthcoming.

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.

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.

Bordalo, P., Coffman, K., Gennaioli, N., and Shleifer, A. (2019). Beliefs about Gender. *American Economic Review*, 109(3):739–773.

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.

Collischon, M. (2019). Is There a Glass Ceiling over Germany? *German Economic Review*, 20(4):e329–e359.

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.

Dale, S. and Krueger, A. (2002). Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables. *Quarterly Journal of Economics*, 117(3):1491–1527.

Fernandes, A., Huber, M., and Vaccaro, G. (2020). Gender differences in wage expectations. Technical Report arXiv:2003.11496, arXiv.org.

Filippin, A. and Ichino, A. (2005). Gender Wage Gap in Expectations and Realizations. *Labour Economics*, 12(1):125–145.

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). Decomposition Methods in Economics. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 4, pages 1–102. Elsevier.

Firpo, S. P., Fortin, N. M., and Lemieux, T. (2018). Decomposing Wage Distributions Using Recentered Influence Function Regressions. *Econometrics*, 6(2).

Francesconi, M. and Parey, M. (2018). Early gender gaps among university graduates. *European Economic Review*, 109:63 – 82.

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.

Kiessling, L., Pinger, P., Seegers, P., and Bergerhoff, J. (2019). Gender Differences in Wage Expectations: Sorting, Children, and Negotiation Styles. CESifo Working Paper No. 7827, CESifo Munich.

Kiviet, J. F. (2016). When Is It Really Justifiable To Ignore Explanatory Variable Endogeneity in a Regression Model? *Economics Letters*, 145:192–195.

Kiviet, J. F. (2020). Testing the impossible: Identifying exclusion restrictions. *Journal of Econometrics*, forthcoming.

Klößner, S. and Pfeifer, G. (2019). The Importance of Tax Adjustments when Evaluating Wage Expectations. *The Scandinavian Journal of Economics*, 121(2):578–605.

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.

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

Osikominu, A., Grossmann, V., and Osterfeld, M. (2020). Sociocultural Background and Choice of STEM Majors at University. *Oxford Economic Papers*, 72:347–369.

PersonalMarkt Services GmbH (no date). `http://www.personalmarkt.de/de/`. Accessed on December 24, 2012.

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.

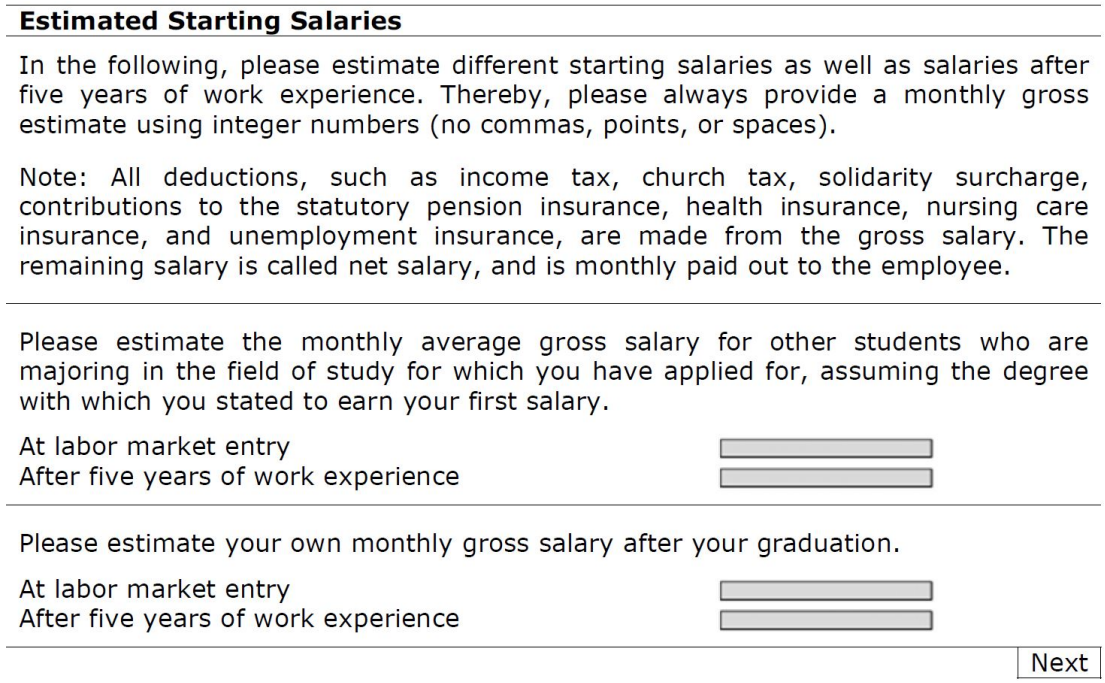
Wiswall, M. and Zafar, B. (2018). Preference for the Workplace, Investment in Human Capital, and Gender. *The Quarterly Journal of Economics*, 133(1):457–507.

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

Appendix

A Survey Questions Eliciting Salary Expectations

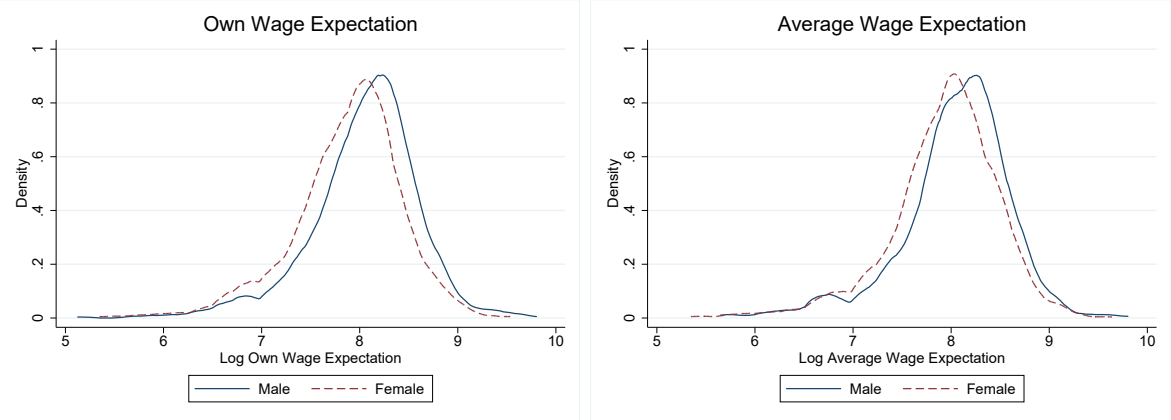
Figure 8: Survey Questions To Elicit Expected Starting Salaries



Notes: The figure shows an extract of the questionnaire the prospective students had to answer during the online survey, which focuses on estimated starting salaries.

B Additional Empirical Evidence

Figure 9: Densities of Expected Starting Salaries



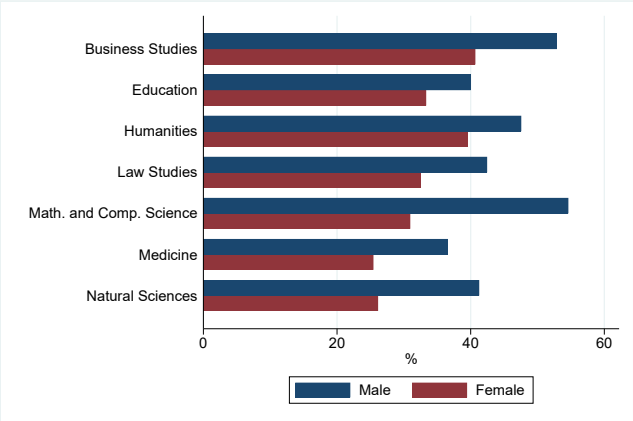
Notes: The left panel shows kernel density estimates of the logarithm of expected own starting salaries. The right panel shows kernel density estimates of the logarithm of expected average starting salaries. In both panels we use an Epanechnikov kernel. *Source:* Student Survey, own calculations.

Table 6: Gender Gap in Expected Average Starting Salary

	(1) Raw	(2) With Controls
Female	-0.1162***	-0.0865***
	(0.0248)	(0.0252)
Field of Study		✓
Intended Degree		✓
Observations	2048	2048

Notes: The dependent variable is the logarithm of the expected average starting salary for other students within same field of study and with the same degree. Column (1) includes only the female dummy variable as regressor. Column (2) additionally includes dummy variables for the field of study and intended degree. Standard errors are shown in parentheses. *, ** and *** indicate statistical significance at the 10%-, 5%-, and 1%-level, respectively. *Source:* Student Survey, own calculations.

Figure 10: Expected Own Starting Salary Above Actual Average Starting Salary



Notes: Figure shows share of students for which the expected own starting salary exceeds actual average starting salaries in their chosen field of study. *Source:* Student Survey and PersonalMarkt Services GmbH (date), own calculations.

C Empirical Support for the Linearity Assumption in the Decomposition Analyses

One potential concern w.r.t the Oaxaca-Blinder decomposition at the mean as well as the decomposition of the unconditional quantiles is that consistent estimation of the explained and the unexplained part requires that the linearity assumption is satisfied in eqs. (2) and (7), respectively (Barsky et al., 2002; Firpo et al., 2011, 2018). If the relationship is nonlinear, the counterfactual salary expectation that women would have, if they formed their expectations in the same way as men, is not equal to $\mathbb{E}[\mathbf{x} \mid G = 1] \cdot \boldsymbol{\beta}_0$ in eq. (3), which decomposes the gender gap in the mean of the log expected salary and the mean of the RIF of the τ-th quantile of the log expected salary, respectively, into the explained and unexplained part. Therefore, as an additional robustness check, we apply the reweighted regression approach described by Firpo et al. (2011). The idea of this approach is to reweigh the observations of male students so as to align the distribution of the characteristics of male students to that of female students. The reweighted regression approach allows us to estimate the specification error arising if the regression model is misspecified. A specification error close to zero indicates that the linear model is accurate. In addition, we use the method to calculate the reweighting error. If the reweighting function is consistently estimated, the reweighting error should be close to zero.

Applying the reweighted regression approach in the decomposition of the gender gap at the mean for the benchmark specification we find a specification error of -0.001, which is not statistically different from zero. Moreover, the reweighting error is very close to zero and statistically insignificant, which indicates that the reweighting factors are consistently estimated.[18] In case of the decomposition of unconditional quantiles, we also find no statistically significant specification error at any decile of the distribution of the expected starting salary. This suggests that the linear specifications seem justified empirically.

[18] The detailed results are available on request.