The Importance of Tax Adjustments when Evaluating Wage Expectations

Stefan Klößner (Saarland University)
Gregor Pfeifer (University of Hohenheim)*

March 20, 2017

Abstract

Using elicited expectations of future gross salaries, we evaluate characteristics causing German students to make larger or smaller estimation errors. While students seem to underestimate actual salaries by 18 percent, we show that these errors are highly attributable to misconceptions of the progressive income tax. Developing a suitable adjustment procedure, we correct students' estimates and find that errors decline by 12 percentage points. Conducting regression analyses, we reveal strong connections with students' age, gender, work experience, secondary school track, and knowledge about student loans. These results change notably if not controlling for students' misconception of the tax system.

Keywords: Wage Differentials; Human Capital; Progressive (Income) Tax; Returns to Education

JEL: H24; I26; J24; J31

*Pfeifer (corresponding author): University of Hohenheim, Department of Economics, 520B, D-70593, Stuttgart, Germany; +4971145922193; g.pfeifer@uni-hohenheim.de. Klößner: Saarland University, Department of Economics, Campus Bldg. C3 1, 66123 Saarbrücken, Germany; s.kloessner@mx.uni-saarland.de. The authors are grateful for valuable remarks from Nadja Dwenger and Aderonke Osikominu as well as participants of WIEM and VfS. Jan Frühwald provided excellent research assistance. We particularly thank Stefan Witte, who contributed to earlier versions of this research. Administrative support by Saarland University's officials is gratefully acknowledged.

1 Introduction

A central assumption of human capital theory is that people select their ideal type and extent of schooling to maximize life-cycle utility—or income. This view is closely linked to the premise that people are able to correctly and rationally forecast income streams of alternative investments in education. Thus, ex ante (perceived) earnings—in contrast to ex post (realized) earnings—affect the extent to which higher education will be pursued. However, it is questionable whether students have an accurate perception of how alternative educational decisions actually influence their prospective income. Indeed, as Lusardi and Mitchell (2014) summarize, many people around the world are financially illiterate, urging to incorporate the fact that financial knowledge is a form of human capital.

Within this framework, we want to shed light on students' expectations regarding future salaries and, hence, evaluate their ability to make accurate predictions of labor market outcomes. Using a rich dataset of more than $2,000$ applicants for Saarland University, Germany, we test whether students have knowledge of their own potential gross salary as well as of salaries for other graduates, on average. We test the accuracy of this knowledge regarding both the student's actually chosen major and for alternative ones. In so doing, we use elicited expectations, examining whether students' estimates are representative for actual salaries observable in the labor market and whether potential estimation errors are driven by specific characteristics.

Our results show that students substantially underestimate actual starting salaries by about 18 percent. However, using specific items from our survey, we can show that estimation errors are highly attributable to students' misconception of the progressive income tax system. We explicitly asked students to estimate net equivalents of given gross salaries in order to test their understanding of the tax code. Corresponding results suggest that students do not perceive the progressivity of income taxation at all, rather expecting a constant or even decreasing tax rate for increasing gross salaries. As a consequence, we adjust students' salary estimates and re-evaluate their knowledge of future labor market outcomes. For this purpose, we provide a stepwise adjustment procedure which first computes perceived net estimates and then re-translates those figures into adjusted, i.e. corrected, gross estimates. Due to the adjustment, students' estimation errors drop by 12 percentage points, showing that applicants do have a quite correct idea about what salaries to expect in the future when tax issues are properly accounted for. Even though applicants' adjusted expectations are quite in line with labor market outcomes, some level of error still remains—6% on average. When conducting regression analyses, we find that the accuracy of students' estimations is particularly driven by age, work experience, the secondary school track, knowledge about student loans, and gender. Notably, we show that using unadjusted expectations of future salaries would yield considerable changes across several covariates in our model, resulting in both economically and statistically different estimates.

While our work can be viewed as an extension to the stream on financial (il)literacy, it is most closely related to a rapidly growing literature concerning elicited expectation measures. Betts (1996), Dominitz and Manski (1996), and Wolter (2000) show that there is substantial variation in students' beliefs about future wages. While such studies reliably demonstrate that students perceive a positive return to college education (also Menon, 2008 and van der Merwe, 2011), such perception varies by field of study and personal traits. Brunello et al. (2004) find that, in addition to these factors, country- and university-specific variables also play a role. Regarding precision, Betts (1996), Jerrim (2011), as well as Botelho and Pinto (2004) state that senior students' estimates are significantly more accurate than those of first-year students, questioning the assumption that students make schooling decisions based on complete information about the labor market. Consequently, this called for studies as conducted by Zafar (2011), analyzing how students form expectations and showing that learning indeed plays a key role within that process. Stinebrickner and Stinebrickner (2012) as well as Stinebrickner and Stinebrickner (2014) also incorporate subjective expectations into models of choice behavior, analyzing the updating of beliefs and its influence on later outcomes. In the same context, Wiswall and Zafar (2015) and Arcidiacono et al. (2012) study specific determinants of college major choice, showing that beliefs about earnings and ability, but also personal tastes, play a significant role. Zafar (2013) adds that non-pecuniary factors might be particularly important for females. Finally, Webbink and Hartog (2004) gauge the accuracy of students' estimates by comparing them with their own (later) realizations, finding that the same variables that influence expectations also have an influence on students' realizations some years later.

The results of these different studies point to the fact that students recognize the positive relationship of earnings and education. However, findings are mixed regarding the accurateness of students' wage perceptions. Moreover, students' estimates seem to vary substantially with respect to their field of study and individual characteristics, but also other important factors that have been ignored for quite some time. While, gradually, researchers account for more soft factors (e.g., taste) in addition to the common estimation error drivers (e.g., school grades), so far, studies completely abstract from potential taxation issues. This is surprising as, first, income taxes clearly belong to the hard factors of pecuniary influences. Second, as long as eight years ago, Lusardi et al. (2009) admonished that even many employees, i.e. people having *finished* their studies, could still not give precise answers regarding their salary in connection with corresponding tax burdens prior to receiving their first pay check.

We contribute to the literature by making a first step towards closing this gap. To the best of our knowledge, we are first to, against the background of elicited expectations, explicitly ask students about their understanding of the tax system. Thereby, we are able to come up with a formal adjustment procedure to construct their true, i.e. tax corrected, wage expectations and corresponding estimation errors. Consequently, this is the first study using tax adjusted estimates to conduct standard inferential analyses examining the drivers of wage expectation errors. We demonstrate that corresponding results significantly differ compared to analogous analyses using unadjusted estimates which would yield biased estimates and, eventually, misleading conclusions regarding the underlying mechanisms. Moreover, we close the gap of missing evidence from Germany on that topic. While students' wage expectations have been analyzed in different settings across various countries, they have not been applied to German cases so far. This country, with a population of about 80 million, basically offers tuition-free education which is utilized by roughly 2.8 million university students of which about 500,000 are freshmen. Since institutions as, e.g. the OECD, keep suggesting to increase enrollment and graduation rates for tertiary education,[1] and also economists as Autor (2014)

[1] For an overview, see `https://www.oecd.org/edu/Education-at-a-Glance-2014.pdf`, last accessed

present evidence on the rising payoff to college education, it becomes more and more important to investigate whether those young people actually are well informed regarding future labor market outcomes and on what characteristics this might depend. Finally, using data on slightly more than $2,000$ students, we draw a considerably larger sample than most of the relevant literature which mostly works with some hundreds of observations. With that rich of a data pool, we are able to precisely analyze the extent to which students' elicited expectations of their future salary depend on different characteristics.

The paper proceeds as follows: Section 2 gives an overview of our sample and shows first descriptive evidence on salary estimates. Section 3 explores students' misconception of the tax system and further provides our adjustment method for correcting (false) salary estimates. In Section 4, we present detailed evidence showing how big of an influence the adjustment has on students' estimates and how they approximately converge to actual job market salaries. Throughout Section 5, we conduct regression analyses examining the influence of students' personal traits on remaining, i.e. adjusted, estimation errors. Moreover, we compare such findings to similarly unadjusted analyses revealing specific differences. Section 6 concludes.

2 Data & Descriptive Evidence

During the enrollment application processes in 2011 and 2012 at Saarland University, Germany, we surveyed students on their beliefs about starting salaries.[2] The survey's URL was e-mailed to all prospective students applying for a university place in these years while only students submitting a complete application within the application deadline were considered. In 2011, 500 students completed the survey; in 2012, $1,561$ students responded. Part of that increase is due to the fact that we were able to add two more subjects of study (Education and Medicine).[3] Throughout Appendix A, we present further details concerning our sample and how it compares to national and international figures, making sure that we can indeed

August 28, 2016.

[2] The first wave of the survey actually took place in 2010 and served to learn about administrative and organizational issues, the survey design, comprehensibility of questions, response rates, etc. Based on these experiences, the survey was slightly edited and officially rolled out in 2011 to gather a two-year data sample.

[3] In 2011, this was not possible due to administrative reasons.

infer valid results from our data.

The survey started with questions regarding the prospective students' field of study. We asked for which degree (Bachelor, Master, State Examination) and for which field of study the student had currently applied for. Students had to state whether they would aspire to obtain an additional degree afterwards (Master, Second State Examination, or a Doctoral Degree), and with which of those degrees they would aim to earn their first salary.

In the second part of the survey, students had to answer two different types of questions about monthly gross salaries. The first one asked students about expectations of their own monthly salary and about their estimates for average others who studied within the same field. In both cases, students referred to the degree with which they intended to earn their first salary. These estimates were reported for salaries at labor market entry and for salaries after five years on the job (hence, four separate questions). From this point onwards, they will be referred to as 'estimates of field related starting salaries for self (for others)' and as 'estimates of field related salaries for self (for others) after five years'. The second type of questions asked students about their estimates of average monthly gross salaries for other students in different fields of study (Business Studies, Education, Humanities, Law Studies, Mathematics/Computer Science, Medicine, and Natural Sciences). In this case, estimates were provided without conditioning on degree (Bachelor, Master, etc.). Again, starting salaries and salaries after five years on the job were reported (hence, 14 separate questions). Note that, before we explicitly asked students for those *gross* salary estimates, we provided detailed explanation on the difference between gross and net salaries.

In the third part of the survey, students had to provide information on their personal and family background as well as on their potential future income and profession. The following characteristics were considered: gender; age; work experience; final grade in secondary school; whether the student's mother or father did study (graduated from college) and, if so, what major discipline; the student's intention concerning living at her parents' house while studying; whether the student expects to receive "BAfoeG"[4] and, if so, how much; the school system

[4] The German *Bundesausbildungsfoerderungsgesetz*—short BAfoeG—is a federal law on support in education providing students from a weaker financial background with funding, specifically, with affordable student loans.

in secondary school;[5] the federal state in which the student obtained her higher education entrance qualification; the importance of an above-average salary; the influence of income expectations on the student's major decision; the student's favorite branch of business and her work experience in this branch. A description of students' overall background characteristics is given in Figure 6 in Appendix B.

Last, students had to estimate net equivalents of given gross salaries (€ 1,500, € 3,000, and € 4,500). These estimates will be used to 1) test and control for the possibility that students have an inaccurate understanding of the income tax system and 2) potentially correct students' salary estimates and re-evaluate their knowledge of future wages. Eventually, for the purpose of comparing students' *estimates* to *actual* figures, we are using data on starting salaries provided by the compensation consultancy PersonalMarkt which offers the largest database of actual salaries for Germany.[6] By using *starting* salaries as our actual benchmark, we implicitly condition on experience level (first-time employee) and broadly also on age (young age cohort, university graduates). This is important since we will later use students' estimates for salaries at *labor market entry* when we examine how corresponding estimation errors depend on certain characteristics and, pivotal for this paper, on the progressive income tax.

Table 1 presents descriptive statistics of students' beliefs about salaries. It is noteworthy that all estimates of salaries after five years are greater than related starting salaries. This finding shows that students plausibly expect salaries to grow with increasing experience, in accordance with, i.a., Dominitz and Manski (1996) and Betts (1996). Concerning *field related salaries*, average differences between expectations for self and average others are nearly absent, both for salaries at labor market entry (€ 2,802 for self and € 2,857 for others) and for salaries after five years (€ 4,033 for self and € 4,002 for others). Regarding estimates of *salaries in different fields*, students estimate the highest salaries for graduates in Medicine and Law Studies, both at labor market entry and after five years. Precisely, the mean estimate of salaries in Medicine (Law Studies) constitutes € 3,251 (€ 3,020) at labor market entry and € 5,076

[5] Until 2011, it was mandatory for students to complete 9 years in secondary school—so-called *G9* high school system—in order to receive their general qualification for university entrance (*Allgemeine Hochschulreife* or *Abitur*). Since then, completing 8 years was sufficient—so-called *G8* high school system.

[6] PersonalMarkt Services GmbH offers a pool of about 1.6 million different salaries covering 260 professions over 60 branches. `http://www.personalmarkt.de/www/uu.home.jsp`, last accessed April 20, 2016.

Table 1: Descriptive Statistics of Estimated Salaries

Estimated Salaries	Obs.	Mean	SD	Min	Median	Max	Skew.	Kurt.
Field Related Starting Salaries								
For Self	2050	2801.69	1186.23	200	2800	14000	1.51	10.13
For Others	2048	2857.31	1171.58	300	2800	14000	1.44	10.31
Field Related Salaries After 5 Years								
For Self	2050	4033.26	1948.87	350	3600	23000	2.72	18.51
For Others	2049	4002.15	1872.74	208.33	3750	27500	3.09	25.48
Starting Salaries in Different Fields								
Business Admin.	1536	2703.79	1073.02	300	2500	15000	2.19	20.09
Education	1517	2539.12	919.58	450	2500	8750	1.15	7.65
Humanities	1491	2221.72	934.54	300	2000	14000	2.48	24.08
Law	1520	3020.25	1221.77	300	3000	16000	1.92	15.19
Math & Comp.Sci.	1493	2942.84	1156.13	400	3000	14000	1.66	13.14
Medicine	1532	3250.68	1462.96	400	3000	20000	2.37	19.26
Natural Sciences	1492	2778.13	1053.42	400	2725	12000	1.25	9.15
Salaries After 5 Years in Different Fields								
Business Admin.	1528	3924.84	1696	700	3700	20000	2.9	22.27
Education	1513	3415.75	1242.82	700	3200	19000	2.79	24.81
Humanities	1487	3099.1	1350.15	700	3000	22000	4.1	44.37
Law	1513	4526.56	2020.87	600	4000	26000	3.05	22.5
Math & Comp.Sci.	1491	4172.58	1784.07	900	4000	25000	3.29	28.95
Medicine	1529	5076.35	2457.87	600	4800	40000	4	41.84
Natural Sciences	1487	3894.63	1602.12	11.67	3570	23000	2.85	25.47

Table 2: Actual Starting Salaries

	Business Studies	Education	Humanities	Law Studies	Math. & Comp.	Medicine	Natural Sciences
Actual	3341.04	3196.96	2559.84	3696.64	3578.96	3766.40	3741.20

Note: Actual starting salaries were obtained from the compensation consultancy PersonalMarkt (`http://www.personalmarkt.de/www/uu.home.jsp`).

(€4,527) after five years on the job. On the contrary, students estimate the lowest salaries for graduates in Humanities and Education. The estimated starting salary in Humanities (Education) is €2,222 (€2,539); the corresponding salary in Humanities (Education) after five years on the job is €3,099 (€3,416).

However, if we compare such *estimated* salaries against *actual* (starting) salaries, displayed in Table 2, it becomes evident that students heavily underestimate future labor market incomes. This difference is huge for, e.g. Medicine (€3,250.68 vs. €3,766.40) or Natural Sciences (€2,778.13 vs. €3,741.20) and at least €-264.64 for Humanities. Over all seven majors, students underestimate future starting salaries by about 18%.

3 Tax Misconception and Adjustment

Before we evaluate the (in)accuracy of students' salary estimates, we turn our focus to a potential source of bias throughout such evaluations. When asking students about their monetary expectations, we requested to provide *gross* estimates.[7] It is not unlikely, however, that young people do not possess proper knowledge of the income tax code, gross-net relations, social contributions, tax rates, tax burdens, etc. This could severely bias wage expectation estimates and hence, succeeding policy implications and recommendations. To the best of our knowledge, none of the related studies so far tries to control for students' understanding of the tax code. In this survey, however, we explicitly asked students about their appreciation of the tax schedule. Thereby, we are able to construct their true wage expectations and corresponding estimation errors.

Relation Between Net and Gross Estimates

As most countries do, Germany applies a progressive income tax system where the tax rate rises as the taxable income increases. Hence, people with more income pay a higher percentage of that income in tax than those with less income. Furthermore, employees have to pay social contributions (statutory pension insurance, health insurance, nursing care insurance, and unemployment insurance); thereby, the level of social contributions levied, again, depends on the gross salary. The tax burden of an employee, thus, results from the payable amount of taxes on her income and from her level of social contributions that she is required to pay. Since employers draw off these quantities from the employee's monthly salary and pass them directly to the responsible agencies, the employee receives a net paycheck.

The statutory net amount is defined by subtracting tax amount (tax), solidarity surcharge

[7] We did ask for gross, not net, since there is no universal net salary—it is always case-sensitive. Moreover, for our comparisons and the calculation of estimation errors, we use *actual* data on salaries, which are always reported as gross figures.

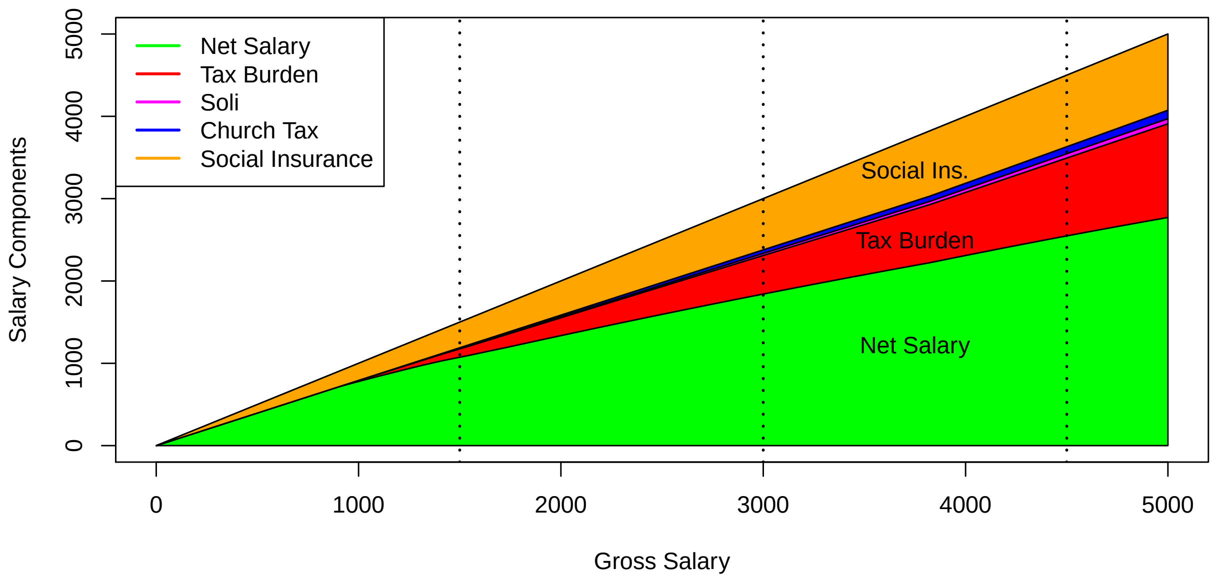


Figure 1: Components of Tax Adjustment of Salary Estimates

(*soli*),[8] church tax (*church*), and social payments (*social*) from the gross salary:

$$\text{gross\_to\_net}_{\text{actual}}(\text{gross}) := \text{gross} - \text{tax}(\text{gross}) - \text{soli}(\text{gross}) - \text{church}(\text{gross}) - \text{social}(\text{gross}) \quad (1)$$

The rules how to determine all these quantities are quite complicated and can be found in Appendix C.[9] In the end, the relation between gross and net salaries as defined in (1) is governed by a concave strictly increasing piecewise quadratic function which, together with the other tax components, is displayed in Figure 1.

In order to survey students' understanding of the progressively rising tax rate, they were asked to provide estimates of net equivalents of the following gross salaries:[10] €1,500, €3,000, and €4,500, which were specifically chosen because the gross-to-net-relation is approximately

[8] To finance the reunification of Germany, a 5.5% solidarity surcharge is levied from all taxpayers on top of income or corporation tax.

[9] We assume the following criteria: payroll year 2012, no children, tax class I, no tax allowances, church tax liability, and statutory health insurance with a contribution rate of 15.5 percent. Such criteria most likely mirror the characteristics of the average, i.e. single and young, university graduate entering the labor market.

[10] Note that these questions were asked on the last page of the survey where maneuvering back to the salary estimates was not an option. Therefore, the construction of the survey ensures that there is no *anchoring effect* which could have influenced students' salary estimates.

Table 3: Comparison of Estimated and Actual Net Salaries

Gross	Est. Net	Est. T.R.	Act. Net	Act. T.R.	Diff. Net	Diff. T.R.
1,500	1,032.07	31.20	1,074.32	28.38	-42.25	2.82
3,000	2,118.42	29.39	1,841.91	38.60	276.51	-9.22
4,500	3,194.70	29.01	2,548.21	43.37	646.50	-14.37

Note: Gross are the three gross salaries for which students were asked to estimate corresponding net values as given in *Est. Net*. *Est. T.R.* is the corresponding tax rate (in percent) calculated as (Gross-Est. Net)/Gross. *Act. Net* gives the actual or true net value reflecting the criteria explained in the text. *Act. T.R.* is the corresponding actual tax rate (in percent). *Diff. Net* provides the difference between *Gross* and *Est. Net*, while *Diff. T.R.* displays the difference between *Est. T.R.* and *Act. T.R.* (in percent). All differences' means are significantly different from 0 at the 1% level.

linear between these incomes, see Figure 1 (vertical, dotted lines). Table 3 compares the corresponding estimates with actual net values showing that the *actual net salaries* of the queried gross salaries are €1,074, €1,842, and €2,548. In turn, this implies corresponding *actual tax rates* (in percent of the gross) of 28%, 39%, and 43%, respectively. However, students' *mean estimates* of the corresponding three net salaries imply an *estimated tax rate* of 31%, 29%, and 29%, respectively. Thus, even though the estimated net equivalent of a gross salary of €1,500 seems to be in rough accordance with the actual net salary, students do not perceive the progressive taxation of income at all. Contrarily, they rather expect a constant or even a slightly decreasing average tax rate for increasing gross salaries, i.e., a regressive tax system.

Adjustment of Salary Estimates

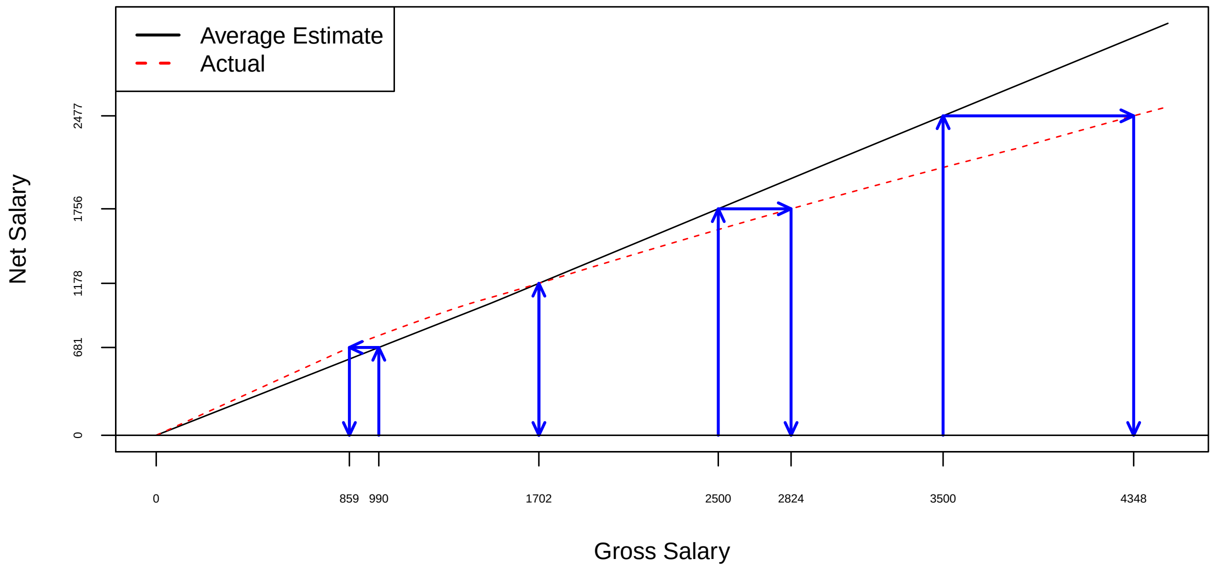
Recall that the survey asked students to report *gross* estimates for expected salaries and that students on average underestimated gross salaries by about 18%. Combining these two facts, underestimation of the tax burden and the gross salaries, it might well be the case that students possess solid knowledge about *net* salaries but fail to express corresponding gross estimates due to their inability to translate net salaries into gross salaries. Put differently, the underestimation of the tax burden presents a potential source of error for students' gross salary estimates. Conceivably, therefore, reported gross salaries are not representative for the actually known and possibly correct net salaries. In order to control for this potential source of error when estimating gross salaries, we adjusted all salary estimates such that gross estimates are representative for students' perceived net salaries. This adjustment is done in several steps. First, for every student, linear interpolation is used to translate her estimates $\text{net}_{1500,i}$, $\text{net}_{3000,i}$, and $\text{net}_{4500,i}$ into a function converting gross incomes to net incomes:

$$\text{gross\_to\_net}_i(\text{gross}) := \begin{cases} \text{net}_{1500,i}\,\dfrac{\text{gross}}{1500} & : \quad \text{gross} \leq 1500 \\ \text{net}_{1500,i} + \dfrac{\text{gross} - 1500}{3000 - 1500}(\text{net}_{3000,i} - \text{net}_{1500,i}) & : \quad 1500 \leq \text{gross} \leq 3000 \\ \text{net}_{3000,i} + \dfrac{\text{gross} - 3000}{4500 - 3000}(\text{net}_{4500,i} - \text{net}_{3000,i}) & : \quad 3000 \leq \text{gross} \end{cases} \tag{2}$$

These student-specific functions are then used to calculate the perceived net salaries by applying gross_to_net_i to the gross salary estimates provided by student i. Finally, the perceived net salaries are converted to adjusted gross salary estimates by using the actual relation between gross and net salaries as defined in (1). Overall, gross salaries were thus adjusted by replacing an estimate gross_i of student i by $\text{gross\_to\_net}^{-1}_{\text{actual}}(\text{gross\_to\_net}_i(\text{gross}_i))$. This is exemplified in Figure 2 which shows the adjustment procedure for the average student. For instance, an estimated gross salary of €990 per month is first translated into a perceived net salary of €681 which is then converted into an adjusted gross salary estimate of €859. As actual and estimated net salary coincide for an estimated salary of €1,702, this estimate remains unchanged while gross estimates of €2,500 and €3,500 are adjusted to €2,824 and €4,348, respectively. Thus, for the average student, small estimates are adjusted downwards due to the overestimation of taxes for small salaries while large estimates are adjusted upwards because of the underestimation of taxes for large salaries.

However, we emphasize that the adjustment was student-specific as Figure 3 reveals: while for some students, adjusted estimates are below supplied estimates (subject 723), they are above supplied estimates for others (subject 1464), and sometimes the adjustment changed estimates only slightly (subject 1283, which is characterized by a rather exact understanding of the gross-net relation). On average, however, adjusted estimates are well above supplied gross estimates, reflecting the average student's overall underestimation of the tax burden.

Adjusted gross estimates are presented in Table 8 in Appendix B, which also compares them to the unadjusted ones as seen in Table 1. Remarkably, each and every mean of the adjusted estimates is distinctly greater than corresponding means of the original estimates.



Note: The blue arrows indicate how supplied (estimated) gross salaries are first converted into net salaries and then into adjusted gross salaries. The black solid line displays the average of the supplied gross to net relations while the red dashed line indicates the true gross to net relation.

Figure 2: Gross-Net Conversion of Salaries: Actual Values vs. Students' Estimates

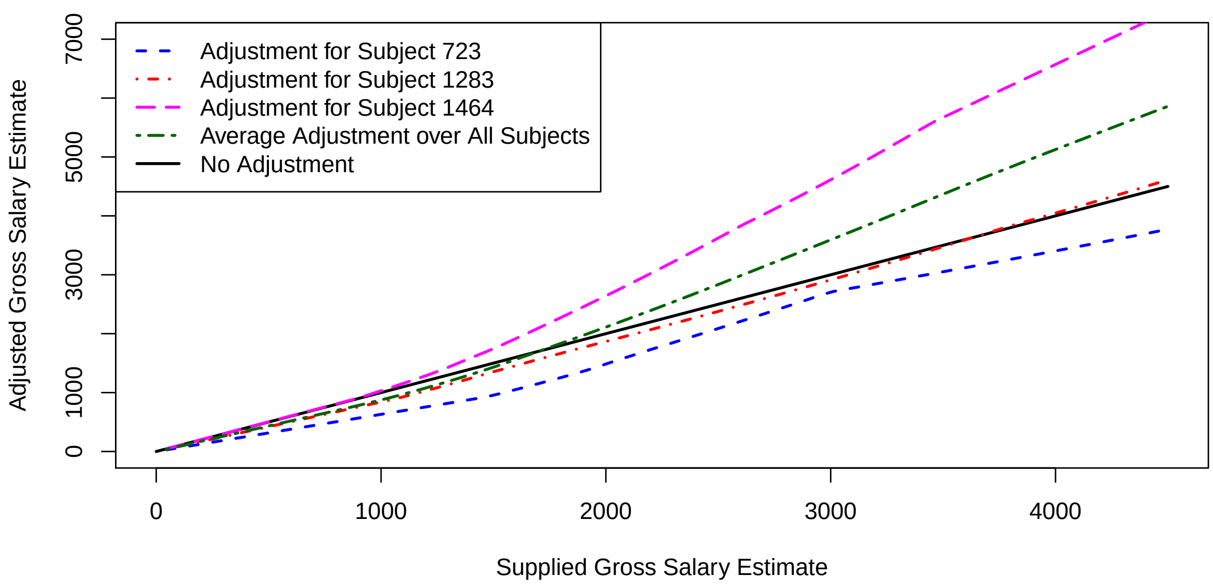


Figure 3: Illustrations of Tax Adjustment for Different Students

Starting salaries, on average, increased by approximately €500, salaries after five years by about €1,000. It seems that, due to a lack of knowledge about the true gross to net calculation, students provided notably too low estimates of gross salaries, which, after our adjustment, not only developed into the "right" direction, but also to a large extent so.

4 Accuracy of Un/Adjusted Estimates

Mean Estimate for Self

In this section, the accuracy of students' estimates of field related starting salaries for *self* is evaluated. Manski (1993) classified expectations for own earnings as conditional since they depend on students' personal characteristics and abilities. Hence, these expectations are subjective and do not (solely) test students' knowledge of the labor market. Put differently, anticipated salaries for themselves might to some extent reflect students' market information, i.e. average observed or anticipated salaries, but *primarily* reflect perceptions on where they end up themselves within the salary distribution: they may consider themselves better or worse than average others observed or anticipated in the labor market. Consequently, hereinafter, estimation errors will not be computed at the *individual* level. However, comparisons between students' *mean* estimates by discipline for which they have applied for and actual, field-specific salaries observed in the labor market are supposed to actually provide meaningful information.

Figure 4 compares actual mean salaries per discipline with students' respective mean expectations. One can easily notice that students' *unadjusted* expectations are considerably lower than actual salaries in each and every discipline, those differences all being statistically significant at the 1 percent level (compare Table 9 in Appendix B). Precisely, the differences in Medicine (€-1,067), Natural Sciences (€-839), Law Studies (€-774), as well as Education (€-731) show the greatest magnitude. The mean of expectations for Humanities is closest to the actual salary with a difference of €-265. In order to see whether the systematic underestimation is (also) due to an inaccurate understanding of the tax system, the previous analysis is conducted again, using *adjusted* salary estimates according to our procedure from Section 3. Again, Figure 4 illustrates corresponding results. The consequent difference is extremely eye-

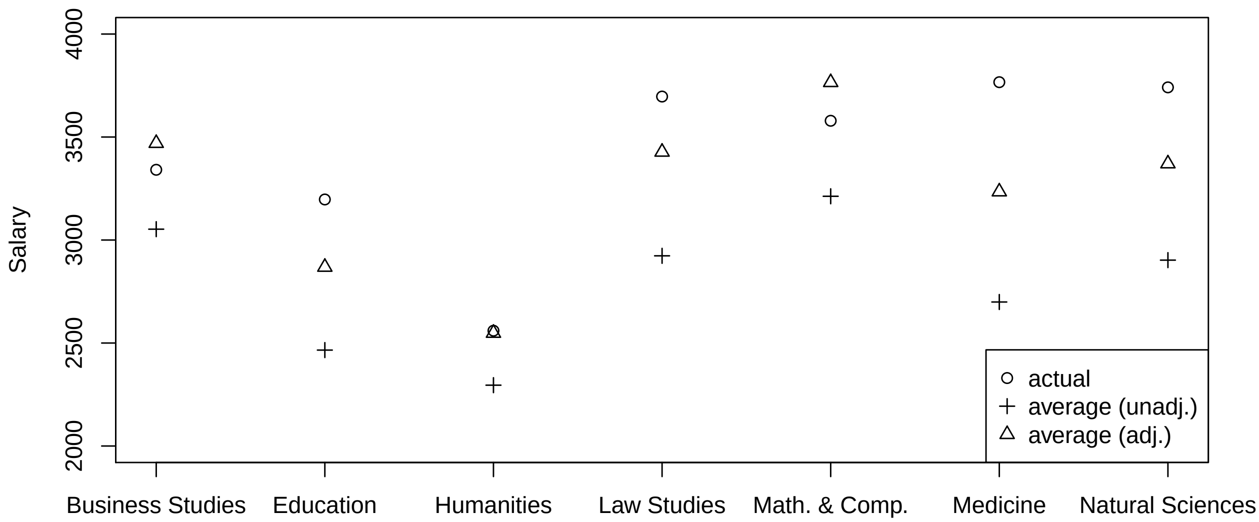


Figure 4: Actual Salaries Compared to the Distribution of the Field Related Starting Salary for Self by Discipline Applied For (Un/Adjusted Estimates)

catching with gaps between the mean of expectations and the corresponding actual salary considerably reducing in every single discipline. In the field of Humanities it even virtually disappears (€-12). Due to such convergence of actual and (adjusted) estimated wages, but also because now differences in expectations for Math/Computer Science and Business Studies are positive (€186, €129) as compared to unadjusted figures (€-367, €-288), a systematic under- or overestimation is no longer observable. However, even though those differences now appear to be notably smaller compared to the unadjusted case, all but two (Math/Computer Science and Humanities) stay statistically significant (compare Table 9 in Appendix B). This means that, even after the tax adjustment procedure, there still remains quite some error in students' beliefs regarding future wages which, on top, (still) appears to be heterogenous within and across different fields of studies.

Starting Salaries in Different Fields

We further examine *errors* in students' estimates of starting salaries in *different fields* of study. The survey asked students to estimate starting salaries in different fields for *average others* rather than for themselves. These perceptions (on others), indeed, reflect market information (Manski, 1993), and, consequently, corresponding salary estimates are useful to test students' knowledge of the overall labor market. Actually, mean signed errors (MSE) of estimates are computed and analyzed, where occurring (substantive) estimation errors would raise doubt on students' information (accuracy) about salaries. In accordance with Wolter (2000), the MSE is defined as the percentage deviation of an estimate from its actual value. It can be formalized as follows:

$$\text{MSE} = \frac{\text{Estimate} - \text{Actual}}{\text{Actual}} * 100 \tag{3}$$

Table 4 presents t-tests for the significance of MSEs across fields of study, i.e. whether the corresponding field specific mean of MSE equals zero. As might be expected based on our previous findings, all MSEs based on *unadjusted* estimates are negative and highly significant. Students made the smallest mistakes when estimating salaries in Humanities (-13.2%) and Medicine (-13.7%), the largest MSEs can be observed when students estimate salaries in Natural Sciences (-25.7%). The overall MSE is -18.6%, which is a considerably larger value than comparable values found by Betts (1996) (-5.8%) and Wolter (2000) (-5.3%).

For *adjusted* estimates, however, absolute values of errors decreased sharply in each field of study. The largest MSE still is observable for estimates in Natural Sciences (-14%); the smallest mistakes stem from students' estimates in Math and Computer Science (-3.4%) and Medicine (3.5%). Noteworthy, the latter is even positive. After the adjustment, the overall MSE drops by about 12 percentage points and, hence, closely converges to corresponding figures found by Betts (1996) and Wolter (2000): -6.1%. Note that, still, every single field shows statistically significant errors, again pointing to the fact that students' knowledge of the overall labor market, especially its salaries, is somewhat flawed.

Next, we go one step further and consider whether students, having applied for different disciplines, differ in their knowledge about salaries *conditional on fields of study*. Figure 5 depicts boxplots of students' overall MSE, broken down by the field they have applied for. Remarkably, all boxes based on *unadjusted* estimates are fully below the value of zero, with the exception of Math/Computer Science; it slightly projects over the zero bar. The equivalent

Table 4: Significance of the Mean Signed Errors in Different Fields (Un/Adjusted Estimates)

	Mean			Standard Error		
	Undaj.	Adj.	Diff.	Undaj.	Adj.	Diff.
Business Studies	-19.07	-7.66	11.42	0.82	1.17	0.56
Education	-20.58	-10.52	10.05	0.74	1.07	0.53
Humanities	-13.21	-5.75	7.46	0.95	1.27	0.56
Law Studies	-18.30	-3.73	14.56	0.85	1.23	0.60
Math. and Comp. Science	-17.77	-3.38	14.40	0.84	1.23	0.60
Medicine	-13.69	3.54	17.23	0.99	1.45	0.69
Natural Sciences	-25.74	-14.02	11.73	0.73	1.06	0.52
Overall	-18.56	-6.13	12.43	0.73	1.04	0.52

Note: Mean gives Mean Signed Errors (as defined in the text) for unadjusted (*Unadj.*) and adjusted (*Adj.*) estimates per field of study, as well as the corresponding difference (*Diff.*). *Standard Error* provides appendant standard errors based on one-sample *t*-tests for the mean.

boxes, using *adjusted* salaries, are also presented in Figure 5. Strikingly, after the adjustment, all boxes considerably jump to the right now covering the value of zero. We can simply visually determine, again, how much estimates improve as soon as we control for students' misconception of the tax system. However, the medians of the MSEs are all (significantly) below zero, still indicating a slight underestimation of future wages.

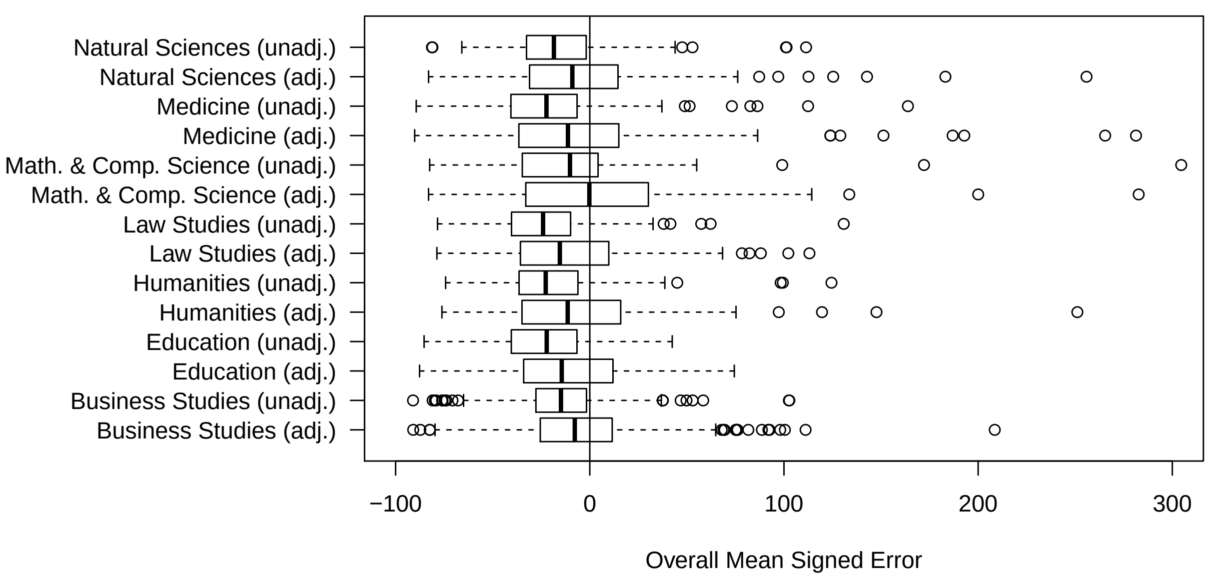


Figure 5: Distribution of the Overall Mean Signed Error by Discipline Applied For (Un/Adjusted Estimates)

Table 5 provides exact numbers of corresponding MSEs in different fields. Recognizably, estimates of students who have applied for Math and Computer Science indicate the smallest overall mistake, followed by students applying for Natural Sciences and Business Studies, while estimates of students applying for Education, Law Studies, Medicine, and Humanities show the largest estimation errors. After the adjustment of salary estimates, this ranking more or less stays the same. However, absolute values of students' MSEs all decrease considerably.

Overall, this evidence is in line with the analysis of students' salary estimates for *self*: students' estimates of salaries in *different fields* are highly attributable to an inaccurate understanding of the interaction of gross and net salaries, i.e., the income tax system. This also and especially holds true if we condition our estimates on respective fields of study. However, after controlling for this tax issue, students still underestimate actual salaries.

Table 5: Mean Signed Errors by Discipline Applied For (Un/Adjusted Estimates)

	Business Studies	Education	Humanities	Law Studies	Math. & Comp.	Medicine	Natural Sciences
Unadjusted	-15.01	-22.98	-20.01	-22.83	-11.15	-21.50	-14.28
Adjusted	-4.97	-10.70	-6.10	-12.14	2.96	-7.02	-0.44

5 Explaining Un/Adjusted Estimation Errors

Regression Model

Even though we could show that, by using a suitable tax adjustment procedure, students' estimation mistakes of future wages declined significantly, there still remains a notable level of error. In a final step, we want to determine the drivers of those residual estimation mistakes. We therefore make use of our rich dataset, conducting regression analyses where we try to examine how the (remaining) variation in students' salary estimates can be explained by personal traits. Moreover, we also want to see how such results change compared to standard analyses not correcting for potential tax issues. Since we observed that students' estimation errors change significantly after our adjustment procedure, we can expect the importance of different error-drivers to change, too.

Considering the literature, we note that using the pure MSE as dependent variable could yield incorrect interpretations, since a positive coefficient of a regressor does not provide information on whether this regressor affects higher positive or smaller negative errors. Therefore, an *absolute* version of the percentage error is used which we compute as follows:

$$\text{Error}_i = \frac{1}{7}\sum_{f=Bus}^{Nat} \log\left(1 + \left|\frac{\text{Estimate}_{i,f} - \text{Actual}_f}{\text{Actual}_f}\right|\right) * 100 \tag{4}$$

We employ the *log* of the error to attenuate large absolute percentage errors. More specifically, we use $\log(1+\ldots)$ instead of $\log(\ldots)$ since we want to avoid that very small absolute percentage errors lead to artificially large negative values.[11] In order to allow for the investigation of student i's general knowledge of salaries in different fields rather than her knowledge of salaries in one specific field, the *overall* log absolute percentage error of different fields is employed as the dependent variable of our regression analysis. Hence, this variable refers to the mean of the log of the (absolute) percentage errors over the seven fields (f runs from *Bus*iness Studies to *Nat*ural Sciences) the student was asked to provide estimates for. Considering only such students who gave estimates for *every* field of study, we estimate the following regression model:

$$Y = \alpha + \beta_0 D + \beta_1 X_1' + \beta_2 X_2' + \beta_3 X_3' + \epsilon \tag{5}$$

where Y denotes the (adjusted) error defined in (4); D is a dummy variable for year 2012; vector X_1' refers to variables capturing the student's personal background (gender, age, work experience, and the final grade of secondary school); vector ${X_2}'$ refers to variables indicating the student's family background (mother/father studied, living at parents' house while studying, receiving BAfoeG, the school system the student comes from, and the federal state where the student went to school); vector X_3' refers to variables that consider the student's future academic and working career (discipline and degree applied for, degree with which student aims to earn first salary, highest targeted degree, the importance of an above-average salary, the influence of income expectations on the student's major decision, and her favorite branch of business); finally, ϵ is a stochastic disturbance term. For simplicity, subscripts referring to time and unit dimensions are suppressed. Several sensitivity checks with respect to this regression model (type of error, scaling, etc.) are discussed throughout Appendix A.

[11] This is different from the literature, where, without any specific reasoning, simply $\log(|\,\text{MSE}\,|)$ is used.

Results

The regression output is presented in Table 6: we consider three different specifications, using either unadjusted (columns one, three, and five) or adjusted expectation errors (columns two, four, and six) as outcome variable. In the first and fourth column, the dependent variable is regressed on variables capturing the students' personal background (X_1') only. In the second and fifth column, variables indicating the students' family background are added (X_2'). Finally, variables capturing the students' academic and working career are included in the third and sixth column (X_3'). All regressions include the year 2012 dummy whose coefficient turns out to be statistically insignificant across all adjusted as well as the second and third unadjusted specifications, mitigating concerns about combining the two waves of data. Moreover, note that the (adjusted) R^2 is totally in line with, or even higher than, comparable values found by, e.g., Betts (1996) or Wolter (2000).

Interestingly, *female* students make significantly larger estimation errors than male students (the reference group). Recall that analyzed estimates in this regression are estimates of salaries for average other students, rather than estimates for self. Therefore, the finding that female students make larger estimation errors shows that female students are less well informed about actual salaries than male students. Remarkably, the tax adjusted coefficients of females substantially decline, by about half, compared to regressions where we would not adjust our estimates. Therefore, female students make larger estimation errors than male students, but these larger errors are, to a large extent, attributable to females' stronger misconception of the gross to net calculation, i.e., the progressive tax system. Correspondingly, in the regressions based on adjusted salaries, the difference between male and female students' overall absolute percentage error is either not at all statistically significant or at a lower level compared to regressions without the adjustment, where we find significance levels of 1 percent throughout all specifications. Therefore, not taking into account the tax misconception of students, one would not only substantially overestimate the amount of female errors but also their statistical importance.

The coefficients of *age* and *work experience* reasonably suggest that older students as well as students who have already worked in their favorite branch of business make smaller estimation errors. Naturally, such students had more time and opportunities to learn about the labor market and corresponding true wages. Consistent with our data, age is modeled as three dummy variables where the middle category (21-25) serves as the reference group. Older students (26-50) make less errors when estimating future wages, whereas younger students (16-20) perform worse. The coefficients on the latter group show up quite stronger in magnitude for the adjusted specifications. In the third specification, not only the corresponding difference grows to a factor of about two, but also the statistical significance level is higher. For the older group, however, adjusted estimates are not only smaller but also statistical significance level is lower compared to unadjusted estimates. Interestingly, the influence of work experience on students' adjusted estimation error is considerably bigger compared to a regression with unadjusted estimates throughout all specifications. In the full specification, this difference grows even to a factor of about 2.5. Moreover, across all adjusted specifications, work experience features very high levels of statistical significance, whereas, for the unadjusted cases, only specification I exhibits statistical significance, at a very low level.

Like Betts (1996) or Webbink and Hartog (2004), we also controlled for *students' grades* serving as a proxy for ability, since more able students might possess better information on salaries than less able students.[12] Note that we used the category "Missing" as our reference group which comprises students who did not (want to) provide a final secondary school grade. We might think about these students as graduating from secondary school with a bad GPA since, apparently, they do not reveal it. The corresponding coefficients of the adjusted estimates, even though not all showing statistical significance, confirm our thoughts: over all specifications, all groups providing a final school grade make less error when estimating future wages compared to students not disclosing their GPA. Again, this is different when using unadjusted estimates; here, besides the fact that we entirely lose statistical significance, the observed sign-pattern across the three specifications does not appear to be sensible at all.

[12] In order to account for state-specific distributions of grades, we used data on the actual federal states' high school GPA distributions of the years 2011/2012, retrieved from `http://www.kmk.org/statistik/weiterfuehrende-links.html`, last accessed July 28, 2016. Using these data, we translated each student's grade into an ordinal variable which states whether the student belonged to the Top 25%, Top 50%, Bottom 50%, or Bottom 25% of high school graduates within the corresponding state.

Table 6: Regressions – Log Absolute % Error of the Overall Starting Salary of Different Fields (Un/Adjusted Estimates)

	Specification I		Specification II		Specification III	
	unadj.	adj.	unadj.	adj.	unadj.	adj.
(Intercept)	20.34***	26.06***	18.91***	25.54***	17.06***	26.3***
	(1.22)	(1.47)	(1.81)	(2.19)	(2.98)	(3.45)
Year 2012	1.29*	0.73	0.49	0.19	1.18	0.11
	(0.76)	(0.86)	(0.77)	(0.88)	(0.88)	(1.04)
Female	2.72***	1.24	2.74***	1.42*	2.79***	1.64**
	(0.67)	(0.76)	(0.66)	(0.76)	(0.69)	(0.79)
Age Old	-1.24	-0.51	-2.4**	-1.84	-2.84**	-2.23*
	(1.14)	(1.29)	(1.13)	(1.3)	(1.18)	(1.33)
Age Young	3.36***	4.61***	3.51***	4.54***	1.8**	3.59***
	(0.77)	(0.84)	(0.81)	(0.91)	(0.9)	(1.03)
Work Experience	-1.53*	-2.66***	-1.14	-2.37**	-1.04	-2.61***
	(0.84)	(0.93)	(0.86)	(0.96)	(0.9)	(1)
Grade Bottom 25%	0.34	-1.05	0.35	-1.05	0.11	-1.24
	(1.32)	(1.55)	(1.34)	(1.58)	(1.34)	(1.6)
Grade Bottom 25%-50%	-0.01	-2.16*	0.21	-2.23*	0.24	-2.03
	(1.12)	(1.26)	(1.15)	(1.29)	(1.14)	(1.29)
Grade Top 25%-50%	1.19	-0.54	0.92	-0.75	1.27	-0.72
	(1.11)	(1.27)	(1.06)	(1.25)	(1.07)	(1.28)
Grade Top 25%	-0.27	-2.15*	-0.29	-2.16*	0.23	-2.01
	(1)	(1.17)	(1.03)	(1.22)	(1.07)	(1.29)
Mother Studied			1.68**	0.55	1.72**	0.56
			(0.84)	(1.01)	(0.84)	(1)
Father Studied			-0.06	1.31	-0.08	1.01
			(0.75)	(0.88)	(0.76)	(0.9)
BAfoeG			-1.4	-3.02**	-0.92	-2.95**
			(1.06)	(1.25)	(1.07)	(1.31)
No BAfoeG			-2.26**	-3.07***	-1.85**	-3.01***
			(0.9)	(1.06)	(0.9)	(1.09)
8-Year Sec. School			0.54	1.83*	0.45	1.81*
			(0.91)	(1.07)	(0.91)	(1.09)
9-Year Sec. School			2.97***	3.56***	2.64***	3.28***
			(0.91)	(1.06)	(0.89)	(1.04)
Living at Parents' House	Not Controlled For		Controlled For		Controlled For	
Federal States	Not Controlled For		Controlled For		Controlled For	
Discipline Applied For	Not Controlled For		Not Controlled For		Controlled For	
Degree Applied For	Not Controlled For		Not Controlled For		Controlled For	
Degree First Salary	Not Controlled For		Not Controlled For		Controlled For	
Degree Aiming At	Not Controlled For		Not Controlled For		Controlled For	
Influence Income Expectation	Not Controlled For		Not Controlled For		Controlled For	
Importance Above Average Salary	Not Controlled For		Not Controlled For		Controlled For	
Branch	Not Controlled For		Not Controlled For		Controlled For	
Observations	1468	1468	1422	1422	1422	1422
F statistic	6.46***	6.6***	4.32***	4.32***	3.65***	2.89***
R^2 in %	3.83	3.91	6.91	6.91	12.82	10.43
Adjusted R^2 in %	3.24	3.32	5.31	5.31	9.31	6.82

Note: Regression results using as dependent variable the mean of the log absolute percentage errors over different fields. Each specification considers errors based on unadjusted (*unadj.*) as well as tax adjusted (*adj.*) salary estimates. Robust standard errors (HC3) in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

The influence of *parents' educational status* on their children could potentially lead them to make smaller mistakes when providing future salary estimates. Here, for regressions using tax adjusted estimates, we cannot confirm this hypothesis finding small and statistically insignificant effects (of positive signs). For regressions without our tax adjustment, however, we counter-intuitively find that mothers' higher education status negatively influences students' knowledge about future wages, with coefficients showing up statistically significant at the 5% level. Additionally, these effects are up to three times larger compared to adjusted ones.[13]

Interestingly, students who know about the loan *BAfoeG*, by stating that they would either receive or not receive it, make (significantly) smaller errors compared to students who "Don't Know" (the reference group). This is plausible since people who know about BAfoeG certainly gave thought to their own and their family's financial situation, meaning that such people actively deal with current and future income situations—as the student and her family have to fulfill certain prerequisites to be eligible for the loan in the first place and the student has to pay back (part of) the BAfoeG-loan later. That is why also perceived future income streams are important to the student ex ante. However, again, differences between tax adjusted and unadjusted estimates emerge: for the "Yes" group, all of the adjusted coefficients turn out to be statistically significant, whereas none of the unadjusted coefficients does. Additionally, their magnitude is also much higher, with a difference of a factor of about three. For the "No" group, all specifications yield statistically significant estimates, yet, tax adjusted figures again turn out to be bigger in magnitude as well as stronger in terms of statistical significance.

Finally, students who completed an *eight-year* or a *nine-year secondary high school* are compared to students from different school systems (reference group "Other") with respect to estimation errors. Only adjusted estimates for the eight-year students show statistically significant results, comprising effects which are about three times greater than their unadjusted counterparts. Nine-year students show a highly significant effect which is quite stable over all specifications, still a little larger for adjusted estimates. It seems that both groups of students make larger errors than people coming from lower school tracks working their way up till

[13] Note that we ignored the group of students answering "Don't Know" to these questions due to its very limited number of observations.

university.[14] This is plausible for the following reasons: “other” students were not allowed to attend high school directly but were initially tracked to lower-ranked schools. However, they are permitted to switch to a higher track if they show very good grade performances. So, it appears reasonable that, since several years already, they really are concerned with their specific future career, working hard to being able to finally enroll into their desired university major. Contrarily, people coming from a G8 or G9 (regular) high school track do not really have to worry about their admission to university.

All in all, we could see that it makes a tremendous difference whether to use tax adjusted estimates or to simply work with students' raw estimates. For almost all variables, the effect magnitude and corresponding levels of statistical significance do change considerably. What is more, for certain variables, also the plain direction of the effect, i.e. the sign, does change. For such covariates, the mere economic story does not appear to be meaningful at all when using unadjusted estimates. In the end, based on tax adjusted figures, what does (not) matter for having a more precise wage perception is quite intuitive: it is not so much about high school grades, where one is coming from or what the intended field of study is,[15] nor the parents' education; what seems to be more important, though, is the student's experience as well as her involvement in pecuniary affairs. For instance, students who are older, having more work know-how, working their way up from a lower school track till university, as well as students knowing about the BAfoeG loan are making considerably smaller mistakes when estimating future wages. However, also gender appears to be a distinct error-driver.

Causality

So far, we provided direct evidence for a distinct underestimation of tax rate progression and showed that correcting perceived salaries for the estimation error in the tax rates substantially reduces the gap between perceived and actual salaries. Hence, our results strongly imply that errors in estimated gross incomes and in estimated tax rates are positively correlated. However, this does not necessarily imply that flawed tax rate perceptions actually *cause* flawed gross

[14] In Germany, students are tracked very early, after grade four.

[15] When checking the influence of the *field applied for* on the error in students' estimates, we do not find systematically deviating effects compared to the reference group of Business Studies.

salary perceptions.

Therefore, in a next step, we alter our regression model to examine how much of the variation in the estimation error is due to students' tax misconception, while controlling for the known set of covariates which we consider to be sufficient to capture any further relevant influences. More precisely, we regress the deviation of estimated gross salaries from actual average gross salaries, i.e. the *unadjusted* error, on a measure of the error of stated income taxes for a certain income level, conditioning on X'_1, X'_2, and X'_3 from (5). The tax-error measure is, similar to our preferred specification's dependent variable as defined in (4), the mean of the log absolute percentage errors w.r.t. students' estimates of net equivalents for given gross salaries[16] :

$$\text{Tax-Error}_i = \frac{1}{3} \sum_{t=1{,}500}^{4{,}500} \log \left(1 + \left| \frac{\text{Estimate}_{i,t} - \text{Actual}_t}{\text{Actual}_t} \right| \right) * 100 \qquad (6)$$

Table 7 provides corresponding results over the three known (unadjusted) specifications. While most of the discussed coefficients remain quite stable compared to our main regression in Section 5, we obtain the expected positive sign of the tax misconception coefficient (*Tax-Error*). Moreover, this coefficient remains extremely robust over the three specifications, ranging between 0.15 and 0.19 and showing statistical significance at the 1% level. Using a simple R^2 comparison with Table 6, this translates into a share of 7 to 30 percent of wage expectation error which can be additionally explained by accounting for students' tax misconception (*Share of Tax in %*).

This finding strongly suggests that errors in estimated gross salaries and in estimated tax rates are not only correlated, but that a flawed (progressive) tax perception, at least to a non-negligible extent, causes flawed gross salary perceptions to rise. This means that we are not only able to stress that students can more accurately predict net than gross salaries, but also that they understate gross salaries largely because they underestimate the importance of the progressive income tax.

[16] Recall that the students were asked to provide net *estimates* for the following gross salaries: €1,500, €3,000, and €4,500. Their *actual* equivalents were €1,074, €1,842, and €2,548, respectively.

Table 7: Regressions – Log Absolute % Error of the Overall Starting Salary of Different Fields (Unadjusted Estimates) and Accuracy of Perceived Income Taxes

	Specification I	Specification II	Specification III
Tax-Error	0.19***	0.17***	0.15***
	(0.03914)	(0.03944)	(0.03996)
Observations	1468	1422	1422
F statistic	8.49***	4.98***	3.9***
R^2 in %	5.51	8.19	13.81
Adj. R^2 in %	4.86	6.55	10.27
Var. due to Tax	1.68	1.28	0.99
Share of Tax in %	30.49	15.63	7.17

Note: Regression results using as dependent variable the mean of the log absolute percentage errors over different fields. Each specification considers errors based on unadjusted salary estimates only and uses the known set of covariates from Table 6 (suppressed here) plus the tax-error/accuracy measure: *Tax-Error* is the mean of the log absolute percentage errors w.r.t. students' estimates of net equivalents for given gross salaries. *Var. due to Tax* is R^2 *in %* minus the respective R^2 *in %* from Table 6. *Share of Tax in %* is the corresponding share of the wage expectation error which can be additionally explained by accounting for the tax error in comparison to the analogous regression without the tax-error control (*Var. due to Tax*/R^2 *in %*). Robust standard errors (HC3) in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

6 Conclusion

We analyze the extent to which students' expectations of their future salaries are precise and whether they depend on different characteristics. We find that, in general, students' salary estimates are heterogeneous and this variation is correlated with personal traits and the chosen subject of study. On average, students substantially underestimate actual starting salaries by about 18 percent. However, a core insight of this paper is that such estimation errors are highly attributable to students' misconception of the income tax system. We offer a specific possibility to correct for the misleading gross-net conversion and, consequently, find applicants to have a quite correct idea about what salaries to expect in the future. Hence, applicants' adjusted expectations are approximately in line with labor market outcomes. However, even after correcting for students' tax misconception, we still observe estimation errors being evident and strongly correlated with personal traits. The bottom line is that students who are older, with more work experience, knowing about specific student-loans, or working their way up from lower school tracks are making considerably smaller mistakes when estimating future wages. These results, however, would change considerably if we did not control for students' misconception of the tax system, reinforcing the importance of tax adjusted estimates.

Along these lines, our findings strongly suggest that students do not have sufficient information about the tax scheme they will face in the near future, which appears to be particularly problematic for female students. It is obvious that for a well-informed educational choice, (partially) based on prospective monetary returns, it is imperative to have a proper understanding of how the tax system will affect one's financial situation once entering the job market. Studies of people's understanding of taxes and corresponding implications on potential savings showed that with just a little extra effort—at very low cost—it is possible to educate people regarding tax issues triggering desirable consequences (see, e.g., Chetty and Saez, 2013 or Heinberg et al., 2014). Analogously, we should think about incorporating seemingly difficult but practically extremely important topics—how taxes work and how they will heavily influence our future income streams—into early school curricula. Finally, future work using elicited measures of wage expectations should carefully consider people's understanding of the tax code since this potential source of error might not only bias inferential results but also subsequent policy implications.

References

Arcidiacono, Peter, V. Joseph Hotz, and Songman Kang, "Modeling College Major Choices Using Elicited Measures of Expectations and Counterfactuals," *Journal of Econometrics*, 2012, *166* (1), 3–16.

Autor, David H., "Skills, Education, and the Rise of Earnings Inequality Among the "Other 99 Percent"," *Science*, 2014, *344* (6186), 843–851.

Betts, Julian R., "What Do Students Know about Wages? Evidence from a Survey of Undergraduates," *The Journal of Human Resources*, 1996, *31* (1), 27–56.

Botelho, Anabela and Ligia Costa Pinto, "Students' expectations of the economic returns to college education: results of a controlled experiment," *Economics of Education Review*, December 2004, *23* (6), 645–653.

Brunello, Giorgio, Claudio Lucifora, and Rudolf Winter-Ebmer, "The Wage Expectations of European Business and Economics Students," *The Journal of Human Resources*, 2004, *39* (4), 1116–1142.

Chetty, Raj and Emmanuel Saez, "Teaching the Tax Code: Earnings Responses to an Experiment with EITC Recipients," *American Economic Journal: Applied Economics*, 2013, *5* (1), 1–31.

Dominitz, Jeff and Charles F. Manski, "Eliciting Student Expectations of the Returns to Schooling," *The Journal of Human Resources*, 1996, *31* (1), 1–26.

Federal Ministry of Finance, "Programmablaufplan für die maschinelle Berechnung der vom Arbeitslohn einzubehaltenden Lohnsteuer, des Solidaritätszuschlags und der Maßstabsteuer für die Kirchenlohnsteuer," https://www.bmf-steuerrechner.de/pruefdaten/pap2012.pdf (last accessed on 2015-08-03) 2012.

Heinberg, Aileen, Angela A. Hung, Arie Kapteyn, Annamaria Lusardi, Anya Savikhin Samek, and Joanne Yoong, "Five Steps to Planning Success. Experimental Evidence from U.S. Households," Working Paper 20203, National Bureau of Economic Research June 2014.

Jerrim, John, "Do UK Higher Education Students Overestimate Their Starting Salary?," *Fiscal Studies*, 2011, *32* (4), 483–509.

Lusardi, Annamaria and Olivia S. Mitchell, "The Economic Importance of Financial Literacy: Theory and Evidence," *Journal of Economic Literature*, March 2014, *52* (1), 5–44.

_ **, Punam Anand Keller, and Adam M. Keller**, "New Ways to Make People Save: A Social Marketing Approach," Working Paper 14715, National Bureau of Economic Research February 2009.

Manski, Charles F., "Adolescent Econometricians: How Do Youth Infer the Returns to Schooling?," in C. Clotfelter and M. Rothschild, eds., *Studies of Supply and Demand in Higher Education*, University of Chicago Press, 1993, pp. 43–60.

Menon, Maria Eliophotou, "Perceived rates of return to higher education: Further evidence from Cyprus," *Economics of Education Review*, February 2008, *27* (1), 39–47.

Stinebrickner, Ralph and Todd R. Stinebrickner, "A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout," *Review of Economic Studies*, 2014, *81* (1), 426–472.

Stinebrickner, Todd and Ralph Stinebrickner, "Learning about Academic Ability and the College Dropout Decision," *Journal of Labor Economics*, 2012, *30* (4), 707–748.

van der Merwe, Alex, "Earnings Expectations of Typical South African University of Technology First-Year Students," *Education Economics*, 2011, *19* (2), 181–198.

Webbink, Dinand and Joop Hartog, "Can Students Predict Starting Salaries? Yes!," *Economics of Education Review*, April 2004, *23* (2), 103–113.

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

Wolter, Stefan C., "Wage Expectations: A Comparison of Swiss and US Students," *Kyklos*, 2000, *53* (1), 51–69.

Zafar, Basit, "How Do College Students Form Expectations?," *Journal of Labor Economics*, 2011, *29* (2), 301–348.

_ , "College Major Choice and the Gender Gap," *Journal of Human Resources*, 2013, *48* (3), 545–595.

A Sensitivity Analysis

Robustness of Preferred Specification

In addition to regressions of the error of the *overall* starting salary of different fields, regressions *pooling* all starting salary questions in different fields were conducted. The difference compared to the regression model above is that, here, each and every single estimation error is used as a separate dependent variable, i.e. we do not average over the seven log absolute errors. This implies that we now allow for differences in knowledge of future wages across respective fields of study. To account for this new dimension, we include a dummy variable for every respective wage question, i.e. a control for the seven different fields the students were asked to provide salary estimates on. Moreover, we also control for the fact that in one out of these seven estimates, the student gives a wage for the field she actually has applied for. In a nutshell, while we gain a lot in terms of precision, i.e. distinctly smaller standard errors across all specifications, the differences in results between adjusted and unadjusted estimation errors remain very similar to those from Table 6 above.

Moreover, regression results also remained stable when we altered the *dependent variable* by first averaging the MSEs for the different fields and only then applying the $\log(1 + ...)$ transformation. In another try to challenge our results, we used *all observations*, i.e. also such students who did provide salary estimates for less than seven fields of study. In so doing, we additionally included a control for the number of missing fields in order to account for possible selection of fields for which the student might possess (better) knowledge regarding future wages. Again, corresponding regression results were extremely close to the ones presented in Table 6, showing that, indeed, our preferred specification occurs to be very robust.

All aforementioned results are available upon request.

Sample

Moreover, one might worry about sample-specific explanations for our findings which we discuss in the following. To begin with, *Saarland University* and its students could be different from the average German university (student) for several reasons. For instance, the university there might be less renowned than other universities, it might attract relatively poor students, or they might simply have a preference for staying in the area. All of this could alternatively explain why expected wages are lower than the German average. Moreover, students at this respective university may have less educated parents than other students in Germany. If parental income serves as the students' reference point, this might also explain their misconceptions about wages and tax rates, especially at the top. Certainly, such potential influences are directly linked to socio-economic and other factors of the region. Hence, it might be that the *Saarland as a region* is different, e.g., poorer or less developed than the rest of Germany.

First, recall that we not only control for the student's personal background (X_1'), but also for several factors indicating the student's family background (X_2') and the student's academic and working career (X_3'). Especially X_2' should already absorb a lot of the issues one might worry about in the context discussed here. For instance, we can use the state of Saarland as the reference group when controlling for the federal state where the respective student obtained her higher education entrance exam. In so doing, we can see that, while over all specifications students from Rhineland-Palatine seem to provide better wage estimates, students from all the other states are not systematically different from the state of Saarland. An F-test of joint significance supports that finding, yielding a p-value of 0.521. Furthermore, when dropping students who received their high school degree in the state of Saarland, we still find our results to be confirmed.[17]

Second, we check potential university and state specifics outside the scope of our dataset. For important economic measures such as GDP per (state) capita, GDP per working hour, GDP per worker, gross value added per worker, gross income per worker, household income per capita, etc., the state of Saarland pretty much fits the average of Germany.[18] As far as the academic environment is concerned, Saarland University does not appear to be a special case either. In the fall/spring term of 2011/2012, the state of Saarland exhibited 2.65 students per (state) capita and 0.50 freshmen per (state) capita, both ranging in the middle of the distribution across the German federal states. Moreover, figures like student/teacher ratio, male/female ratio, student age distribution, distribution of graduates across fields, distribution of gender across fields, number of exams passed, grades, duration of studies, etc., all more or less range around the German average, not pointing towards peculiarities of Saarland University and its students.[19] Eventually, in a multinational ranking, Saarland University does not rank particularly high or low across a number of items such as Teaching/Learning, Research, Knowledge Transfer, International Orientation, or Regional Engagement.[20]

Summing up, this evidence strongly suggests that our main findings are not sensitive to specifics regarding our sample, i.e., we can safely state to work with and infer results from an overall valid dataset.

[17] We also run our preferred regression for other subsamples, e.g., dropping the youngest age group or the group of students having no working experience. Again, for the remaining coefficients, our results from Section 5 stay strongly robust. All results are available upon request.

[18] For an overview of these figures, see, e.g., `http://www.vgrdl.de/VGRdL/`, last accessed March 14, 2017.

[19] See `https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/BildungForschungKultur/Hochschulen/Hochschulen.html`, last accessed July 27, 2016.

[20] See `http://www.umultirank.org/`, last accessed July 17, 2016.

B Tables and Figures

Table 8: Means of Estimated Salaries (Un/Adjusted Estimates)

Estimated Salaries	Adjusted	Unadjusted	Difference
Field Related Starting Salaries			
For Self	3253.3	2801.69	451.61
For Others	3345.16	2857.31	487.85
Field Related Salaries After 5 Years			
For Self	5027	4033.26	993.74
For Others	5009.28	4002.15	1007.12
Starting Salaries in Different Fields			
Business Admin.	3085.25	2703.79	381.47
Education	2860.52	2539.12	321.4
Humanities	2412.57	2221.72	190.84
Law	3558.63	3020.25	538.39
Math & Comp.Sci.	3458.05	2942.84	515.21
Medicine	3899.8	3250.68	649.12
Natural Sciences	3216.8	2778.13	438.67
Salaries After 5 Years in Different Fields			
Business Admin.	4846.77	3924.84	921.93
Education	4132.42	3415.75	716.68
Humanities	3663.05	3099.1	563.95
Law	5742.11	4526.56	1215.55
Math & Comp.Sci.	5247.58	4172.58	1075
Medicine	6516.06	5076.35	1439.71
Natural Sciences	4837.3	3894.63	942.67

Table 9: Field Related Starting Salaries – Estimates (Un/Adjusted) and Actual by Discipline Applied For

	Unadjusted Estimates			Adjusted Estimates	
	Diff.	Avg.	Actual	Avg.	Diff.
Business Studies	-288.27^{***}	3052.73	3341	3469.84	128.84^{*}
Education	-731.27^{***}	2465.73	3197	2868.58	-328.42^{***}
Humanities	-264.64^{***}	2295.36	2560	2548.00	-12.00
Law Studies	-773.71^{***}	2923.29	3697	3427.33	-269.67^{**}
Math. & Comp.	-366.65^{***}	3212.35	3579	3765.32	186.32
Medicine	-1066.83^{***}	2699.17	3766	3234.51	-531.49^{***}
Natural Sciences	-838.57^{***}	2902.43	3741	3370.52	-370.48^{**}

Note: Actual provides the true wage in a given field of study. *Avg.* equals the respective mean wage expectation. *Diff.* equals Avg. minus Actual. $H_0: Diff. = 0$. $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$.

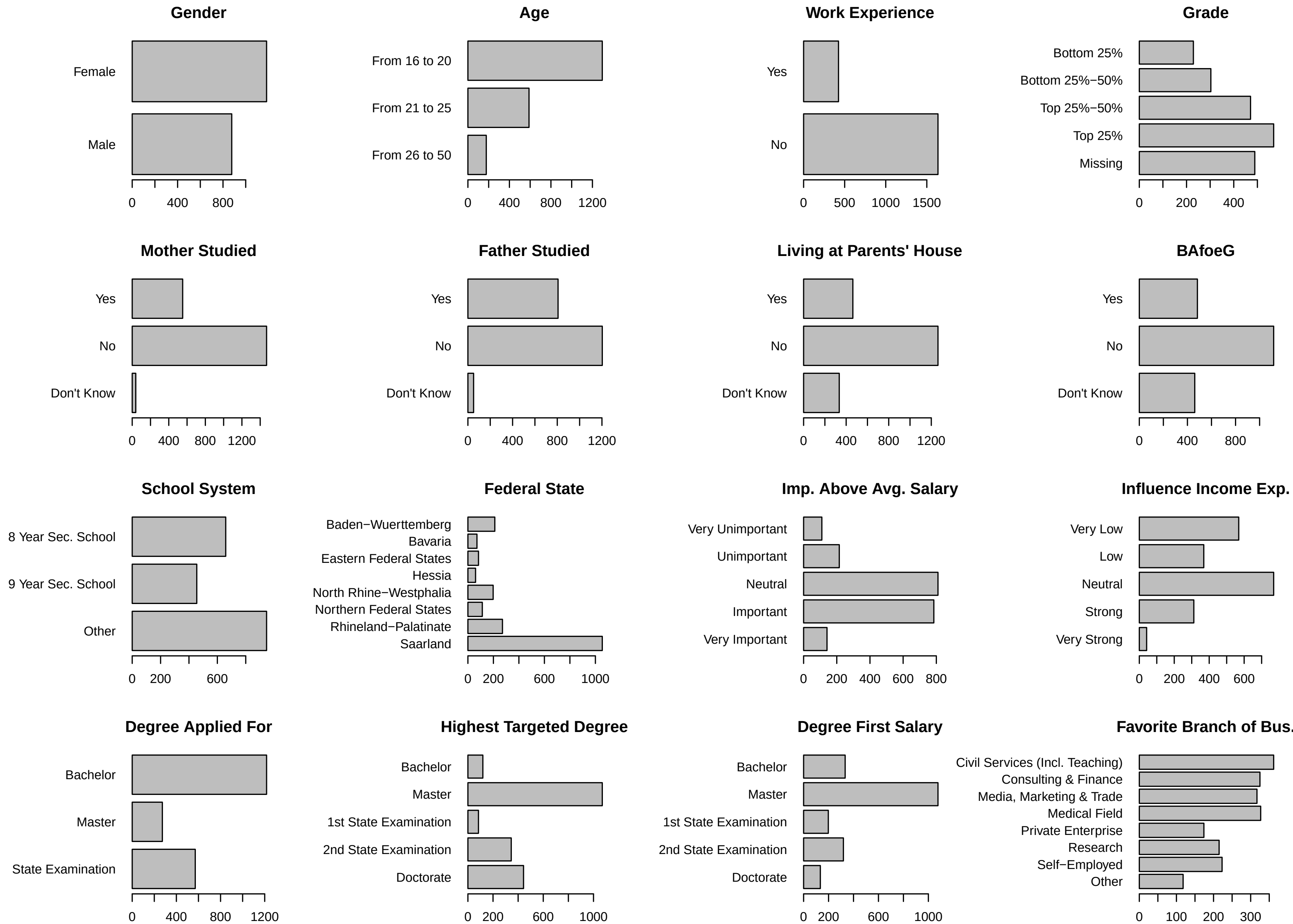


Figure 6: Data Set: Barplots

C The German Tax System

In Germany, the statutory net salary is calculated as $\text{gross\_to\_net}_{\text{actual}}(\text{gross}) := \text{gross} - \text{tax}(\text{gross}) - \text{soli}(\text{gross}) - \text{church}(\text{gross}) - \text{social}(\text{gross})$, where tax denotes tax liability, soli stands for solidarity surcharge, church is church tax (if applicable), and social are statutory contributions to social insurances. The latter consist of unemployment insurance, health insurance, nursing care insurance, and pension insurance. The corresponding formulas are[21]

$$\text{unempl}(\text{gross}) := \begin{cases} 0.015 * \text{gross} & : \text{gross} \leq 5600 \\ 84 & : 5600 \leq \text{gross} \end{cases}$$

for unemployment insurance,

$$\text{health}(\text{gross}) := \begin{cases} 0.082 * \text{gross} & : \text{gross} \leq 3825 \\ 313.65 & : 3825 \leq \text{gross} \end{cases}$$

for health insurance,

$$\text{nursing}(\text{gross}) := \begin{cases} 0.01225 * \text{gross} & : \text{gross} \leq 3825 \\ 46.85625 & : 3825 \leq \text{gross} \end{cases}$$

for nursing care insurance, and

$$\text{pension}(\text{gross}) := \begin{cases} 0.098 * \text{gross} & : \text{gross} \leq 5600 \\ 548.8 & : 5600 \leq \text{gross} \end{cases}$$

for pension insurance. Taken together, these four components make up the mandatory social distributions:

$$\text{social}(\text{gross}) = \begin{cases} 0.20725 * \text{gross} & : \text{gross} \leq 3825 \\ 360.50625 + 0.113 * \text{gross} & : 3825 \leq \text{gross} \leq 5600 \\ 993.30625 & : 5600 \leq \text{gross} \end{cases}$$

Calculating the tax amount which in turn is the basis for both solidarity surcharge and church tax, is done in several steps: first, the monthly gross income gross is converted into an annual income by $\text{gross}_{\text{year}} := 12 * \text{gross}$. The annual income is then used to determine the

[21] Formulas stem from the German Federal Ministry of Finance (2012).

taxable income:

$$
\text{taxable}(\text{gross}_{\text{year}}) = \begin{cases} -36 - 0.16704 * \text{gross}_{\text{year}} & : \quad x \leq 1000 \\ -1036 + 0.83296 * \text{gross}_{\text{year}} & : \quad 1000 \leq x \leq \dfrac{47500}{3} \\ -2936 + 0.95296 * \text{gross}_{\text{year}} & : \quad \dfrac{47500}{3} \leq x \leq \dfrac{7625000}{365} \\ -1036 + 0.86171 * \text{gross}_{\text{year}} & : \quad \dfrac{7625000}{365} \leq x \leq 45900 \\ -5224.375 + 0.95296 * \text{gross}_{\text{year}} & : \quad 45900 \leq x \leq 67200 \\ -8385.463 + \text{gross}_{\text{year}} & : \quad 67200 \leq x \end{cases}
$$

The taxable income then determines the (annual) tax amount, with $x :=$ taxable:

$$
\text{tax}_{\text{year}}(x) = \begin{cases} 0 & : \quad x \leq 8004 \\ \left(669.899132 + \dfrac{x}{10000}\right) * \dfrac{x - 8004}{10000} & : \quad 8004 \leq x \leq 13469 \\ \left(2088.910094 + \dfrac{x}{10000} * 228.74\right) * \dfrac{x - 13469}{10000} + 1038 & : \quad 13469 \leq x \leq 52881 \\ 0.42 * x - 8172 & : \quad 52881 \leq x \leq 250730 \\ 0.45 * x - 15694 & : \quad 250730 \leq x \end{cases}
$$

Thus, the annual tax amount is a piecewise quadratic function of the taxable income. By dividing by 12, we finally get the monthly tax amount: tax $:= \text{tax}_{\text{year}}/12$. While church tax is simply 9% of tax liability (church $:= 0.09 * \text{tax}$), the solidarity surcharge is given by:

$$
\text{soli}(\text{tax}) := \begin{cases} 0 & : \quad \text{tax} \leq 972 \\ \dfrac{\text{tax} - 972}{5} & : \quad 972 \leq \text{tax} \leq \dfrac{38880}{29} \\ 0.055 * \text{tax} & : \quad \dfrac{38880}{29} \leq \text{tax} \end{cases}
$$

In the end, all functions considered above are piecewise linear or piecewise quadratic functions, they are displayed in Figure 1. In particular, the net salary is a concave strictly increasing piecewise quadratic function of the gross salary.