MKIK GVI Kutatási Füzetek

HCCI INSTITUTE FOR ECONOMIC AND ENTERPRISE RESEARCH

MKIK Gazdaság- és Vállalkozáskutató Intézet

Do young humanities and arts graduates have a labor market disadvantage in Hungary?

Data & estimations based on data of the Hungarian Graduate Career Tracking System

2015/5

Do young humanities and arts graduates have a labor market disadvantage in Hungary?

Data & estimations based on data of the Hungarian Graduate Career Tracking System

Hátrányban vannak-e a fiatal bölcsészek a magyar munkaerőpiacon?

Adatok és becslések a Diplomás Pályakövetési Rendszer 2013-as adatfelvétele alapján

Budapest, August 2015

The Institute for Economic and Enterprise Research operated by the Hungarian Chamber of Commerce and Industry (MKIK GVI) is a non-profit, economic research institute indulging in applied research in several subfields of economics.

MKIK GVI **Institute for Economic and Enterprise Research**
Hungarian Chamber of Commerce and Industry

Do young humanities and arts graduates have a labor market disadvantage in Hungary?
Data & estimations based on data of the Hungarian Graduate Career Tracking System

Hátrányban vannak-e a fiatal bölcsészek a magyar munkaerő-piacon?
A foglalkoztatottság és a bérek vizsgálata a Diplomás Pályakövetési Rendszer 2013-as adatfelvétele alapján

MKIK GVI Kutatási Füzetek 2015/5

The study is written by:

Bence Szabó, analyst, MKIK GVI
szabo.bence@gvi.hu
Istvan Janos Tóth, (senior research fellow, MTA KRTK KTI, director, MKIK GVI)
e-mail: *tothij@econ.core.hu*

MKIK Gazdaság- és Vállalkozáskutató Intézet
Budapest

Date: August 30th, 2015

Address: MKIK GVI
1034 Budapest, Bécsi út 120.
Tel: 235-05-84
Fax: 235-07-13
e-mail: *gvi@gvi.hu*
Internet: *http://www.gvi.hu*

Contents

List of Tables:

List of Figures:

Abstract

The study examines the employment status (i) and wages (ii) of young humanities and arts graduates compared to other graduates in Hungary using the 2013 data of the Hungarian Graduate Career Tracking System. The dataset has several limitations regarding reliability and validity, which we attempted to correct to improve the quality of our estimations. The results requires further research. We find that amongst males the conditional overall employment differences for humanities and arts graduates without a degree requirement are not statistically significant. As for jobs requiring tertiary education and salaries disadvantage is present only compared to engineering, information technology and economics or business graduates. Amongst females we find that the conditional employment differences vary greatly by the definition of degree requirement, so our results are unequivocal. However salary disadvantage is statistically significant compared to engineering, information technology and economics or business graduates. Our results suggest that in Hungary young graduates with these three degrees have more advantageous labor market outcomes than others, but the situation of humanities and arts graduates does not differ greatly from the rest of the disciplines.

Absztrakt

A tanulmány a fiatal magyar bölcsészdiplomások elhelyezkedési (i) és kereseti (ii) helyzetét vizsgálja, összehasonlítva más diplomásokkal a magyar Diplomás Pályakövetési Rendszer 2013-as adatait felhasználva. Az adatbázis számos megbízhatósági és érvényességi nehézséggel küzd, amit igyekeztünk korrigálni. A kapott eredmények alapján további vizsgálatok szükségesek. Férfiak esetében azt találtuk, hogy - más tényezők hatásától tisztított foglalkoztatottságot figyelembe véve - nincs statisztikailag szignifikáns különbség a bölcsészek és nem bölcsészek elhelyezkedési esélyei között, míg a kereset esetén a diplomát igénylő állásokat figyelembe véve a különbség csak a mérnöki, informatikai és közgazdász vagy üzleti diplomák esetében van jelen. Nők esetében a foglalkoztatottsági különbségek eredményei jelentősen függnek a diplomát igénylő munkahely definíciójától, így az eredmények nem egyértelműek. Keresetek tekintetében azt találtuk, hogy szintén csak a mérnöki, informatikai és közgazdász vagy üzleti diplomákkal szemben vannak hátrányban a bölcsészek. Az eredményeink azt sugallják, hogy Magyarországon e három utóbbi diploma fiatal végzettjei rendelkeznek a többi szaknál előnyösebb munkaerő-piaci helyzettel, azonban a bölcsészek helyzete nem különbözik jelentősen a többi szakterületétől.

JEL Classification: I21, I26, J44

Keywords: salary differences, employment differences, graduates, humanities and arts graduates, Hungary

Introduction

In this paper we are going to examine the early career stage employment and wage differences of humanities and art graduates compared to their fellow graduates from other fields based on the 2013 database of the Hungarian Graduate Career Tracking System (GCTS, Nyüsti–Veroszta, 2014) provided to MKIK GVI by the Educatio Public Services Non-profit LLC. This topic is relevant in Hungarian policy discussion as well as public discussion about the labor market opportunities of young humanities and arts students: can they find jobs or jobs that require a degree with more or less success than their peers? Do they earn less? If yes: how much less? The answers for this question may be affected to the present Hungarian educational policy.

The Hungarian Graduate Career Tracking System (GCTS) targets the bachelor and master graduates of 32 Hungarian universities with an overall response rate of 16.3%. It has several serious limitations regarding validity and reliability. The full sample is not a random sample and there are no thorough studies regarding sample errors, external supervision of the institutions' practices and their effects on the quality of the survey. However for the target population GCTS is still the best option, with weighting and sample restrictions we intended to create a dataset that is better suited for statistically more reliable estimations. Considering the potential biases and all the weaknesses of the data we can make only cautious observations.

The results suggest that most of the employment and wage differences of humanities and arts graduates compared to other graduates originates from the advantageous situation of engineering, information technology and economics/business graduates while in most cases compared to the rest of non-humanities and arts graduates the differences are not statistically significant and small in scale. For Hungarian graduates the real story appears to be in fact the high return on having a degree from engineering, information technology and economics/business while other fields are lagging behind together.

The structure of the paper is the following. First we are going to look at the deficiencies of the database, how we created the dataset and some basic descriptive statistics. Then we look at the employment gap from different perspectives: overall employment and having jobs requiring a tertiary education. Afterwards we are going to look at salary differences, we check for the robustness of the findings, discuss the results and conclude.

Brief review of the literature on humanities and arts graduates in the labor market

The labor market outcomes of graduates of different study fields have been examined in the past for many different countries, we only refer to some of the most recent studies relevant to our topic. Varga (2013) referred to several international and also Hungarian studies regarding the labor market success of graduates with different fields. All of them suggest that there are great differences in labor market success by fields of degree. The results usually show that graduates with engineering and economics or business degrees are the most successful while graduates with teaching, humanities and arts degrees are amongst the least successful.

For Hungary Varga examined the employment and salary differences amongst graduates using previous GCTS-databases. Using propensity score matching methods Varga (2010) found significant negative coefficients in the case of humanities graduates when using nearest neighbor matching (they earn around 10% less and have 8.4 pp lower employment) and no significant coefficients using stratification matching. She also emphasizes that we should be cautious due to selection bias, which is also very important in our own inquiry. Varga (2013) used the 2011 database of GCTS and found that humanities and arts, natural sciences and social sciences fields have a lower than average rate while teaching, law, information technology, and medicine have a higher than average employment rate. She also examined the salary gap using OLS and quantile regression models. She found that compared to law graduates humanities and arts students earn around 15% less without controls and 8% less controlling for various factors, at other points of the distribution there was no significant coefficient. Quantile regressions can handle the salary differences within fields while OLS only examines average effects, using this method she found that humanities and arts graduates earn 13% less at the 75th percentile and 10% less at the 90th percentile compared to law graduates. Economics/business and information technology graduates have a salary advantage at every point of the distribution, while engineering at the 10th, 25th and 50th percentile compared to the reference group law graduates.

A blog entry by János Köllő (2015) based on the data of the Hungarian Central Statistical Office's Labor Force Survey[1] found no significant employment rate difference between people with engineering and humanities degrees. Based on the Unified Job Classification System (FEOR) he also found no evidence that humanities graduates are employed less in jobs requiring a degree: around 71% of both groups are employed as employees requiring a degree

[1] Ábrahám–Kézdi (2000), appendix: „*The LFS is a quarterly conducted rolling household panel survey of more than 20 thousand non-institutional households. Each time, one sixth of the sample is replaced, so one household stays for 6 quarters in the sample. One main objective of the survey is to provide accurate, ILO-standard measures of participation and unemployment. It also contains rather detailed information about the employment: sector, occupation, and hours worked is provided. It is a household survey, so all information come from the respondents*."

or as managers for 2011-2013 amongst 24-62 year olds. Altogether 79% of humanities graduates and 83% of engineers work but the difference disappears when controlling for gender. He found no difference neither regarding unemployment nor social aid. The working paper of Hajdu–Hermann–Horn–Kertesi–Kézdi–Köllő–Varga (2015) does not find any significant difference in the 30–34 age group either except for the fraction of maternity leave (humanities and arts advantage) and those that retired early (engineer advantage). However cohorts under 30 years are not examined.

Altogether we can summarize that people with humanities and arts degrees are usually found to have lower salaries and a bit lower employment rate amongst young graduates according to GCTS databases, while studies based upon the Labor Force Survey finds no significant disadvantage in employment differences. Also previous GCTS-based findings suggests that altogether economics/business, information technology and engineering graduates have the most success in the Hungarian labor market regarding employment and salary.

About the dataset

We use the 2013 institutional database of the Hungarian Graduate Career Tracking System (GCTS) of the Educatio Public Services Non-profit LLC[2]. The data are based on mostly online, self-report survey results conducted by 32 Hungarian universities amongst graduates of years 2008, 2010 and 2012, so the base population consists of graduates who have received their degrees 5/3/1 years ago respectively. The average response rate was at 16.3% from a population of 148 548: so the original dataset consists of 24 233 observations (Veroszta, 2014).

The database has many basic issues that we have to be aware of before statistical analysis. DPR Monitoring 2014 (Garai, 2014) provides information on how the institutions conducted the surveys. The questioning period usually lasted from May to June. One third of the institutions sent out reminding e-mails twice, four universities sent out only one, the rest more. Nine institutions used other means to popularize GCTS. Most of the institutions signaled that response rates are very low and might cause bias for in-depth analysis: therefore we have to check some important controls, too. During the process the universities attempted to contact all of their graduates mostly by e-mail or phone: the institutions' response rates differ as their base population is obviously different and also because they use different techniques and incentives to encourage responses. They used amongst others motivating letters, reminders, gifts. We do not know however that in practice how often possible mistakes like filling out the survey more than one time or manipulated response rates happened, in theory institutions

[2] We are grateful to Zsuzsanna Veroszta for providing the MKIK GVI the data and answering our questions about the database.

cater to these issues. However some anecdotal concerns remain about the supervision of universities[3].

There are many sources for systematic bias. According to Garai (2014) monitoring revealed that many questions are not sensible or relevant for graduates working abroad and entrepreneurs. According to Berde (2010) the non-reachable graduates can be both from the most and least successful sets of graduates, resulting in unpredictable biases, however most graduates probably refuse to respond consciously. Altogether it is more probable that graduates who are more positively attached to their institutions are overrepresented which might result in positive bias regarding labor market outcomes, too. This can however increase or decrease the estimated gap between humanities and non-humanities graduates depending on their salary and employment distributions. Weighting created by Educatio Public Services Non-profit LLC helps correcting for certain biases: the dataset is weighted so that it is representative of the base population's university, field, year of graduation, gender and schedule (Veroszta, 2014). The results regarding many more questions can be found at Veroszta (2014), this paper covers only the descriptive statistics that are relevant to our topic and we also exclude observations for the purpose of a better sample.[4]

The survey is a dominantly online survey with a response rate of 16.3%, we briefly examine whether this appears to be a good response rate. T. H. Shih and X. Fan (2009) reviewed 35 study results that compared e-mail and mail response rates and found the average to be 33% in case of e-mails. According to them college or higher education populations have a higher response rate on average, and they might have less or no technical problems on average with online or internet surveying causing the overall advantage of mail surveying to diminish for their population type. This argument supports that the use of e-mail surveys for the GCTS is not a problem. The survey provider FluidSurveys found that their average response rate excluding small scale surveys and outliers was 24.8% (Penwarden, 2014). Another survey provider SurveyGizmo (2015) reports an average of 30–40% response rates for internal and 10–15% for external surveys. In MKIK GVI's own experience about online response rates of Hungarian firms this number is around 8-10%. GCTS database seem to have a bit lower response rate than the average taking the fact into consideration that the universities have a chance to build up a connection with their students or former students and that it surveys highly educated people, however it is still higher than expected relative to external and Hungarian online surveys.

[3] We crossed upon anecdotal evidence that in the case of a university there is no unique ID at the start, also meaning in theory that there is no barrier for other graduates to complete the survey, or completing it twice by mistake.

[4] Details about the Hungarian Graduate Career Tracking System can be found in English at: http://www.felvi.hu/pub_bin/dload/DPR_tanulmanyok/dpr_integration_of_data_2013_en_VEGLEGES_web.pdf

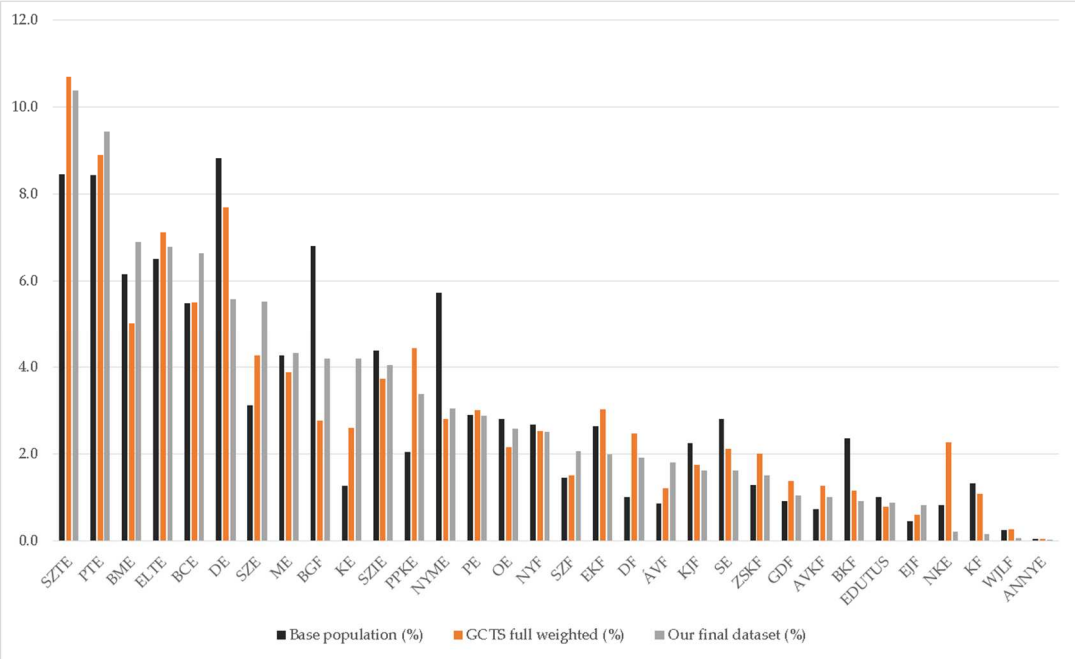
Response rate differences might cause selection bias by gender, income status and ability and therefore influence our results. Micklewright, Schnepf and Skinner (2010) examined the response rate biases of English PISA samples and found that boys, children from low-income families, children with lower school performance are usually underrepresented. They also emphasize that using bigger samples the role of biases increases. This could mean for instance that in the case of engineers the estimated outcomes would be positively biased as their population consists of more men than the average.

Altogether the overall validity and reliability of the data can be questioned: universities have different capabilities to track their graduates, they use different methods and incentives, they only reach a fraction of their former students, the sampling is not random, we have no solid information regarding sampling errors and found no thorough analysis about these questions except for the monitoring document. However this is the best option to study the target group as there are no other greater scale survey specifically on graduates in Hungary. With weighting and restricting the sample on the group that it covers the most reliably we intend to build a dataset for a somewhat reliable statistical analysis.

Figure 1 displays the histogram of the base population, the full dataset with weights on and our final weighted sample of 6670 graduates by university. We collected university base population numbers for the 2013 dataset from felvi.hu[5], and if it was not available, we used the most recent years' data (2012 data for BKF and ANNYE, 2011 data for BGF and NYME). It appears that response rates are very different by universities, the distribution is not similar, and we cannot state anything sure about the potential biases. The imputed base populations seem to be greater than the one they created the weights for, and the calculable base population numbers differ from the numbers provided at the detailed web site (it adds up to 169 060 vs. 147 862 stated by Veroszta, 2014). This proves that it is important to be extremely cautious about the interpretation of our results as the data cannot guarantee anything about the exact measures of the respondents' net salaries or employment data and that it is necessary to draw conclusions only for those that could be cleared of the most obvious biases.

[5] This was suggested by Zsuzsanna Veroszta.
URL: http://www.felvi.hu/felsooktatasimuhely/dpr/eredmenyek/palyakovetesi_adatok_2013.

Figure 1: Distribution of graduates by universities for the base population, the original weighted dataset and the final weighted sample, %



With sample restrictions we intended to create a dataset that focuses on young graduates that are past their early labor market transition phase and can be compared reasonably. We excluded those who failed to provide information on their gender (dropped 1231 observations) or their field of study (dropped 848 obs.) as these were crucial in our estimations. We dropped those also at the start who received their degrees from fields of law enforcement or military (dropped 496 obs.).We also dropped those who graduated in 2012 (dropped 8851 obs.), leaving only the graduates of 2008 and 2010: with that step we intended to "leave time" for graduates to match up with their jobs properly.

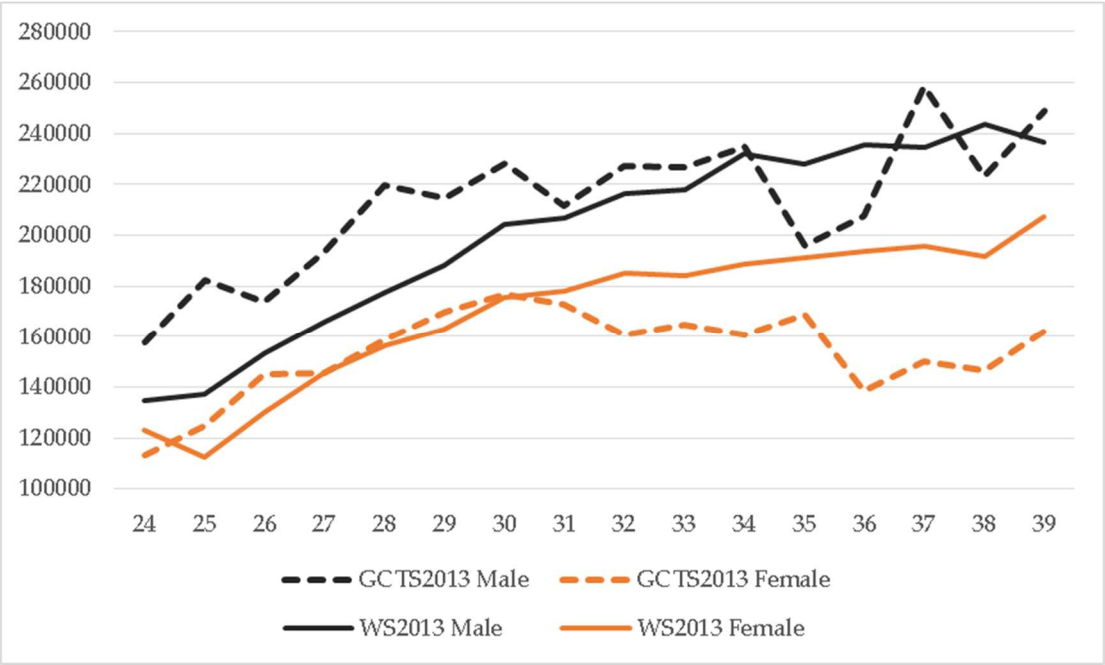
We restricted our sample to the age group of 25–30 year old graduates. This decision was based on looking at the differences between average net salaries of graduates employed in Hungary of the GCTS and the gross graduate salary data of the 2013 Wage Surveys of the National Labor Center[6] (WS, Bértarifa in Hungarian) salaries survey. In GCTS data net salaries are provided by respondents while NLC WS gross salaries are collected from public and private organizations on employees. We calculated the 51% of gross salaries to get around the value of net salaries as it was the average salary tax wedge in Hungary in 2013 according to OECD[7], the numbers naturally are not exact but help use to examine the quality of our data. Figure 2 shows the average salaries on age by gender without young graduates working abroad from

[6] Ábrahám–Kézdi (2000) appendix: *„The WS are very large regular (from 1995 on yearly) cross-sectional surveys of individual earnings and occupational data, also including gender, date of birth, and schooling information. The data are collected from the employers on an administrative basis. They represent the population of firms employing more than 20 persons (more than 10 from 1995)."*

[7] http://www.oecd.org/hungary/taxingwages-hungary.htm

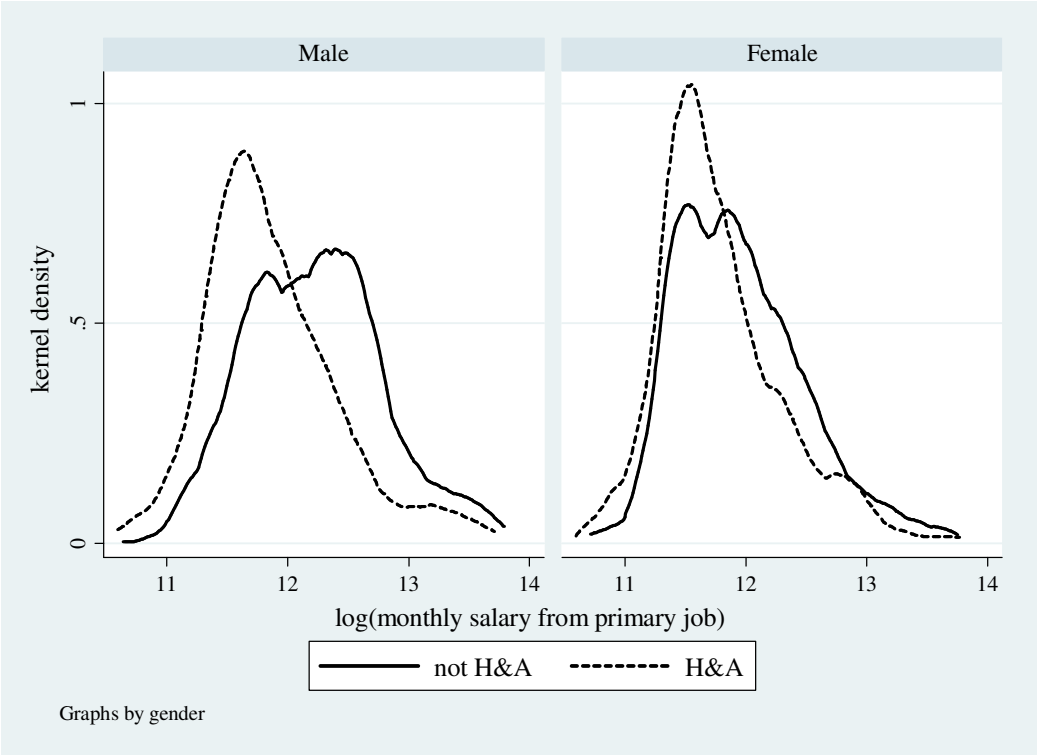
GCTS and only the middle 90% of the distribution of WS for 24-39 year olds. The figure suggests that GCTS data might only be reliable within the 25-30 age interval therefore we decided to go on further with the 7530 observations of only these cohorts.

Figure 2: Wage Survey approximates of graduate net salaries (middle 90%, both genders) and GCTS net salaries (both genders, working in Hungary) for the 24-39 age cohorts , HUF



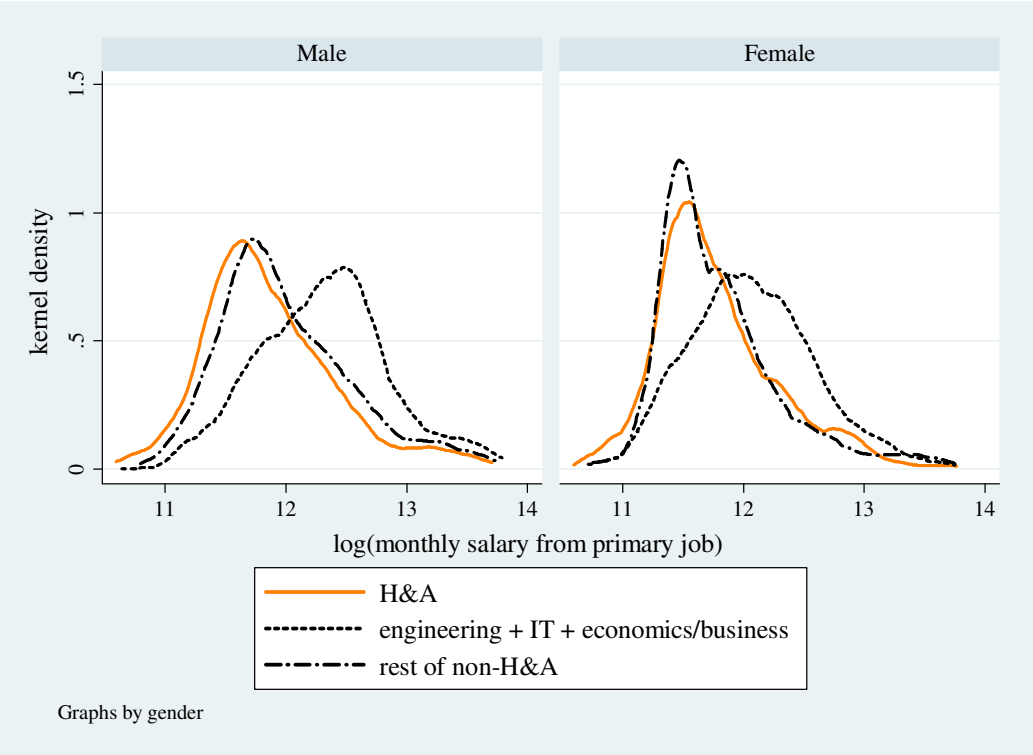
Our key variable is the study field of humanities and art graduates but following Köllő (2015) we included not strictly humanities and art students, but also: political science, cultural anthropology, social studies, social work, psychology and sociology to reflect a bit more on the general everyday use of the phrase. Further on we are going to use H&A as an abbreviation to signal these fields as our group of interest. However if we take a look at the kernel densities of the salaries for the created two groups (Figure 3) we can observe that the non-H&A group seem to hide at least two significantly different groups of people.

Figure 3: Kernel density functions of log(salaries) for humanities and arts graduates and the others by gender



Following Varga (2010) and Varga (2013) we suspect that people with degrees in engineering, information technology or economics/business studies have higher salaries. Figure 4 displays the monthly salary kernel densities for H&A, engineering with IT and economics/business together, and the rest separately. It clearly suggests that in raw terms most of the salary disadvantage of H&A graduates can be attributed to these three professions, and the salary gap seems to be between engineering, information technology or economics/business studies (further on I use the abbreviation E&IT&EB) and the rest of the graduates. The kernel density functions of the H&A and the rest of non-H&A groups also seem to be remarkably similar. Therefore it is advisable to look at the three groups separately.

Figure 4: Kernel density functions of log(salaries) for humanities and arts graduates, engineers, IT and economist/business graduates and the rest by gender



We also exclude graduates that are still full-time students, at home with maternity leave or not participating for other reasons so our final sample only consist of employed and unemployed graduates. With regards to the labor force situation and the different graduate groups Table 1 shows the weighted number of observations and its fractions amongst the group, observation numbers are always rounded to integers. The fraction of employees amongst H&A graduates is 74.5% which is 11,0 pp less than amongst E&IT&EB and only a bit less than amongst rest of non-H&A (1,1 pp). There are also more self-employed (3.7% vs. 1.4% vs. 2.2%) and full-time students (6.3% vs. 3.8% vs. 6.2%) amongst the H&A than the two others respectively. From the point of people on maternity leaves and housewives or other non-participants H&A graduates represent the middle ground. The two groups have the same fraction of housewives (0.6%), H&A group has less entrepreneurs (0.9% vs. 1.7% for both the others) and missing answers (0.0 vs. 0.5% vs. 1.4%) amongst them. However there is a two times bigger fraction of unemployed amongst H&A graduates (8.0%) than the two other groups (3.8% and 4.1% respectively). Further on we are going to examine only the graduates that are employees, self-employed, entrepreneurs or unemployed. It means that altogether 6670 observations stay in the sample.

Table 1: Weighted numbers of observation and frequencies for labor force status within groups of study fields. head and %

	H&A	E&IT&EB	Rest of non-H&A	Total
Employee	1010	3171	1861	6042
%	74.5	85.4	75.6	80.2
Self-employed	50	51	55	156
%	3.7	1.4	2.2	2.1
Entrepreneur	12	62	43	117
%	0.9	1.7	1.7	1.6
Unemployed	108	142	101	351
%	8.0	3.8	4.1	4.7
Full-time student	86	100	152	339
%	6.3	2.7	6.2	4.5
Maternity leave	81	154	194	429
%	6.0	4.1	7.9	5.7
Housewife/other not participating	8	14	22	44
%	0.6	0.4	0.9	0.6
missing	0	18	34	52
%	0.0	0.5	1.4	0.7
Total	1356	3711	2463	7530
	100.0	100.0	100.0	100.0

What kind of bias might this decision result in? It depends on whether the different groups have different proportions regarding full-time students who choose to study further on to avoid unemployment. If a given group has a greater fraction of them it might result in a slightly negative bias regarding employment gap compared to the other groups. In the case of salaries the bias might be the following: we can only observe the salaries of the employed although we might be more interested in the expected value of salaries or salary offers. We can suspect that those who are full-time students to avoid unemployment would receive a lower salary so we probably underestimate the salary gap. However it might also be true that it is simply more common to study further on amongst non-E&IT&EB as these professions have different study structures.

After the sample restrictions the distribution of fields of study can be observed in Figure 5. With weights on the fraction of H&A graduates is at 17.7%, including also other professions that we have already mentioned. For the non-H&A group it shows that most of them are economics or business graduates with the second most frequent group being engineering, so altogether the E&IT&EB-group gives the 51.4% of the sample. The rest is naturally 30.9%,

consisting of the following fields: agriculture, law, medicine, teaching, sports, nature science and partially social science (except for political science, social anthropology etc.).

Figure 5: Fractions of fields of studies for the final dataset (N = 6670)

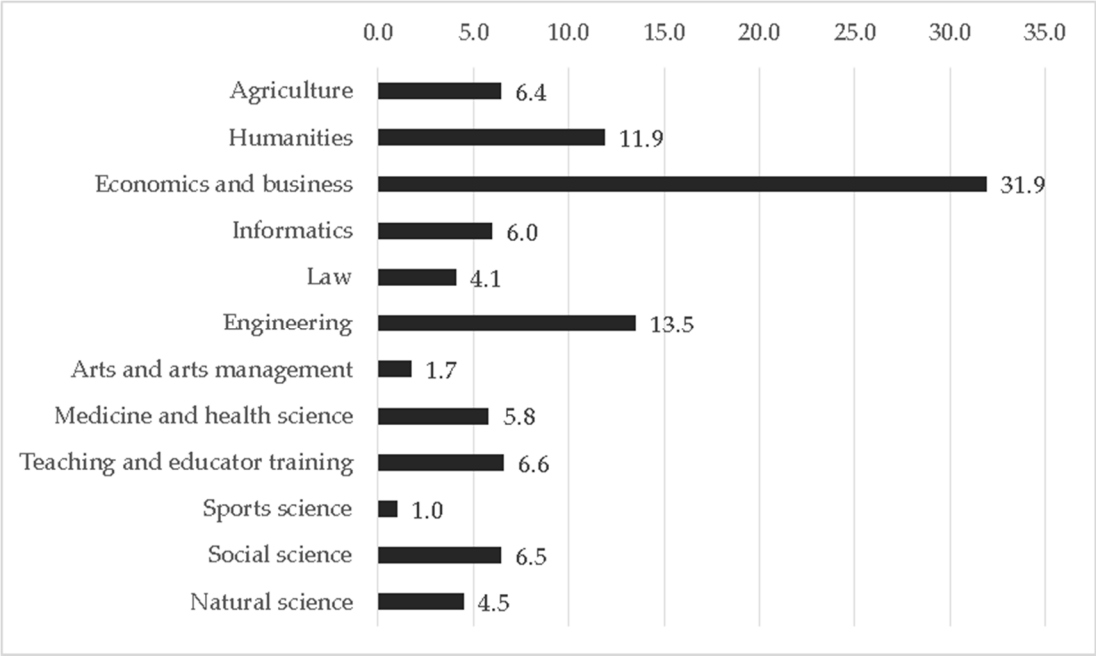


Table 2 describes some differences in characteristics regarding the original and final datasets with weights on. We obviously decreased the mean and standard deviation of age. Concerning gender differences the two datasets are similar. However the distribution by fields of study differs greatly: it seems that for the original dataset consists of a smaller fraction for both H&A and E&IT&EB graduates, while the "rest of" group represents a much greater fraction. Regarding salaries it also seems that the excluded group has a smaller average salary and a lower response rate, which suggests that the more successful group of people remained in the sample. We believe that sample restrictions helped us to focus on the target population, but it is important to show that from various aspects the datasets differ greatly.

Table 2: Weighted mean and standard deviation differences between the original and our final dataset according to main sample features

	Original dataset			Final dataset			Rest (weighted)		
	mean	sd	N	mean	sd	N	mean	sd	N
fraction of females	59.1%	49.2%	24233	59.3%	49.1%	6670	59.0%	49.2%	17563
age	30.7	7.3	22419	27.7	1.5	6670	32.2	8.4	15749
H&A	9.4%	29.2%	24233	17.7%	38.2%	6670	5.9%	23.6%	17563
E&IT&EB	23.3%	42.3%	24233	51.4%	50.0%	6670	11.3%	31.7%	17563
Rest of non-H&A	67.2%	46.9%	24233	30.9%	46.2%	6670	82.8%	37.7%	17563
monthly average wage (HUF)	185,219	125,262	16147	200,166	139,895	5,378	176,758	115,316	10,769

Variables and basic descriptive statistics

In this chapter we introduce the key dependent variables and report the descriptive statistics for some crucial control variables that might inform use about the quality our dataset and also about the basic distributions regarding the H&A and the other graduates. We already defined this variable: we included humanities and art students along with political science, cultural anthropology, social studies, social work, psychology and sociology.

We introduce two types of employment variables: employed in general and employed in a job requiring a degree, the latter having three different definitions. The first definition is based on the question of "what level of education fits your job the most accordingly", where graduates could choose from the following list: PhD; other post-gradual; university/MA/MSc; college/Ba/BSc or it does not demand tertiary education. Those who chose that latest would not count as employed according to this definition. The second and the third definitions are based on the Unified Job Classification System (FEOR-groups)[8]. The third definition includes jobs in management or jobs with requirement of applying tertiary education individually (FEOR main groups 1 and 2), while the second adds to this group "other" jobs requiring tertiary or secondary level education belong in this group (so FEOR main groups 1, 2 and 3). The third definition follows Köllő (2015) but the second one also considers the fact that 25–30 year old employees might still work only as assistants and technicians of a job actually needing a degree.

In the original dataset there are two variables of salaries in thousand HUFs: net monthly salary from the primary job reported by the respondent and also a secondary net salary in case the graduate has a second job. As there were not any major advantages from adding secondary salaries we only used primary net salaries. We used the variable's natural logarithms in the regressions. Of course there are differences in the amount of hours a person worked, the dataset provides us only categories of working hours, but by using them we can control for basic differences in working hours. Table 2 shows that fraction of salaries missing amongst employed is around 16.6% amongst H&A graduates, 14.2% in the case of E&IT&EB and 15.7% amongst the rest, so the difference regarding missing salaries is not remarkable. The selection bias caused by missing offering salaries of the unemployed and the missing salaries are to be treated by Heckman's two-stage estimation procedure in the robustness check, however bias does not seem to be crucial for this sample.

[8] The entire list of FEOR-categories in Hungarian can be found at: https://www.ksh.hu/docs/szolgaltatasok/hun/feor08/feorlista.html

Table 3: Weighted observation numbers and frequencies of missing salaries within groups of fields of study, head and %

	salaries not missing	salaries missing	Total
H&A	896	178	1,074
%	83.5	16.6	100.0
E&IT&EB	2,821	467	3,288
%	85.8	14.2	100.0
Rest of non-H&A	1654	307	1,961
%	84.3	15.7	100.0
Total	5,371	952	6,323
%	85.0	15.1	100.0

The raw differences naturally can originate from the groups' different gender constitution (for instance obviously maternity leave). Table 3 shows that the fraction of women is highest amongst the H&A graduates (75.5%), the second highest amongst the "rest of non-H&A" category and the lowest amongst E&IT&EB (49.8%). This might be a great contributor to the salary and employment differences that everyday people can observe. In general gender influences many other factors that should be controlled for in our analysis and for this reason we are going to analyze the two genders separately.

Table 4: Weighted observation numbers and fractions of genders within groups of fields of study, head and %

	Male	Female	Total
H&A	289	892	1,181
%	24.5	75.5	100.0
E&IT&EB	1,722	1,706	3,427
%	50.2	49.8	100.0
Rest of non-H&A	703	1,358	2,061
%	34.1	65.9	100.0
Total	2,714	3,956	6,670
%	40.7	59.3	100.0

First we take a look at the raw employment differences between the three groups for men and women separately. Regarding overall employment Table 4 shows the rates and weighted observation numbers: both men and women in every category are more than 90% employed, and the H&A group has a small, around 4-5 pp deficit compared to both other groups. These findings accord to Köllő (2015) and Hajdu–Hermann–Horn–Kertesi–Kézdi–Köllő–Varga (2015) about H&A not having a major overall employment gap. Also we can observe an around 1% cross-gender raw employment gap in every category.

Table 5: Weighted observation numbers and fractions of employed by groups of fields of study and gender, head and %

	Employment		
	Male	Female	Total
H&A	91.8%	90.5%	90.8%
	316	976	1,291
E&IT&EB	96.2%	95.5%	95.9%
	1,882	1,864	3,746
Rest of non-H&A	95.8%	94.7%	95.1%
	769	1484	2253
Total %	95.6%	94.1%	94.7%
N	2,967	4,324	7,291

Concerning having a job with a degree requirement Table 5 shows the differences between H&A, E&IT&EB and the rest. First we look at men. From the graduates' answers it seems that amongst the H&A graduates 74.1% work at a place where a degree is needed, and it means an around 16 pp disadvantage vs. E&IT&EB and a 9 pp deficit vs. the rest of non-H&A. If our definition is based on categories of the jobs and includes assistants and technicians it seems that the gap decreases to around 11 pp and 1 pp respectively (85.3% in the case of H&A, that is the most inclusive definition), but without assistants and technicians it stays around 16 pp and 10 pp at a different level (61.2 in the case of H&A, it is the strictest definition). That means a higher employment of assistants and technicians amongst H&A than the other groups amongst males.

Women's situation is a bit different. The H&A group has a disadvantage according to the self-judged definition of around 9 pp against E&IT&EB and an around 10 pp against the rest (level of H&A is 73.8%). Using the second definition the differences are almost entirely diminished: against E&IT&EB it is around 3 pp and around 6 pp against the rest (level of H&A is 84.5%). By the third definition female H&A graduates actually have an around 4–5 pp advantage over the E&IT&EB group but have an astoundingly great 17 pp deficit against the rest of non-H&A. Also the rate for the third definition is only at 55.1% for H&A, while 72.1% for the rest of non-H&A women (higher than men!). Gender gap widens in this category for H&A and E&IT&EB, being extremely high for E&IT&EB at 26%, so amongst them it is more common for women than men to stay at an assistant or technician "other" job level rather than receiving a job requiring a degree and applying tertiary level skill set.

Table 6: Weighted observation numbers and fractions of people having a job requiring a degree by groups of fields of study and gender, head and %

	Employment requiring degree according to the graduate			Employment requiring degree by job FEOR, incl. assistants & technicians			Employment requiring degree by job FEOR, *not* incl. assistants & technicians		
	Male	Female	Total	Male	Female	Total	Male	Female	Total
H&A	74.1%	73.6%	73.8%	85.3%	84.5%	84.7%	61.2%	55.1%	56.5%
	279	867	1145	252	826	1078	252	826	1,078
E&IT&EB	90.2%	82.3%	86.3%	96.7%	87.6%	92.2%	77.1%	50.6%	63.8%
	1,800	1,765	3,565	1,648	1,652	3,300	1,648	1,652	3,300
Rest of non-H&A	82.9%	84.0%	83.6%	86.5%	90.1%	88.8%	71.4%	72.1%	71.9%
	731	1,377	2,109	675	1,324	1,998	675	1,324	1,998
Total	86.7%	81.0%	83.4%	92.9%	87.8%	89.9%	74.0%	59.1%	65.1%
	2,810	4,009	6,819	2574	3,801	6,376	2,574	3,801	6,376

We already looked at the figure for the kernel densities of the three groups' salaries (Figure 4), Table 6 looks at the average monthly net salary differences by gender and field group. The table reveals significant differences across genders and fields regarding the primary salaries. Amongst men H&A graduates earn around 90 thousand HUFs less than E&IT&EB and around 25 thousand HUFs less than the rest of non-H&A graduates (around 35% and 13% respectively) amongst women it is around 50 thousand HUFs and only 5 thousand HUFs (around 24% and 3% respectively). Amongst males there is a quite big employment and salary advantage for E&IT&EB and a smaller one for the rest of non-H&A, too, while for women the employment advantage varies depending on the definition however the wage advantage seems significant only for the E&IT&EB group.

Table 7: Weighted observation numbers and the averages of monthly salaries from primary job by groups of fields of study and gender, HUF

	Primary job monthly net salary (HUF)		
	Male	Female	Total
H&A	170,336	152,339	156,805
	243	736	979
E&IT&EB	260,324	201,423	231,246
	1,560	1,521	3,082
Rest of non-H&A	195,584	157,530	170,820
	631	1176	1807
Total	234,561	175,867	200,218
	2,434	3,433	5,868

There are many other variables that affect employment and salaries which are going to be controlled for in the regressions. We are only going to show two other control variables of

which might have a great influence: categories of working hours in the primary job and position at workplace. The category of working hours means the last average work week for the graduate. Table 7 displays the weighted observation numbers and fractions within the graduate group by categories of working hours for the employed. The dominant group naturally works 30-49 hours, however H&A have a bigger fraction working less than30 hours and a smaller fraction working more than 50 hours which must result in lower averages.

Table 8: Weighted observation numbers and fractions of people in the category of working hours within groups of fields of study

	H&A	E&IT&EB	Rest of non-H&A	Total
less than 30 hours	142	99	168	409
%	13.2	3.0	8.6	6.5
30-49 hours	857	2,915	1,604	5,375
%	79.8	88.7	81.8	85.0
more than 50 hours	52	231	156	438
%	4.8	7.0	8.0	6.9
missing	24	43	33	100
%	2.2	1.3	1.7	1.6
Total	1,074	3,288	1,961	6,323
	100.0	100.0	100.0	100.0

Table 8 shows the average net salaries and weighted frequencies of the previously featured groups. H&A graduates have a salary disadvantage in every category, a relatively big one against E&IT&EB graduates, and a smaller one against the rest of non-H&A, in the dominant 30-49 hours group the difference is only around 8 thousand HUFs. It can also be observed that with increasing working hours there is a 50 thousand HUF increase in salaries for the H&A, 75 thousand HUF increase for the E&IT&EB and a 40 and 75 thousand HUF increase for the rest It suggests that salaries / hours measures probably stay around the same for H&A graduates with increasing working hours along with E&IT&EB graduates while for the rest it does not stand.

Table 9: Weighted observation numbers and the averages of monthly salaries from primary job by groups of fields of study and categories of working hours

	H&A	E&IT&EB	Rest of non-H&A	Total
less than 30 hours	109,474	152,697	128,303	127,739
%	121	85	153	359
30-49 hours	160,292	228,205	168,257	199,778
%	799	2,774	1,501	5,074
more than 50 hours	213,111	302,479	241,439	269,161
%	53	211	147	411
missing	150,629	237,547	170,564	197,981
%	7	12	6	24
Total	156,805	231,246	170,820	200,218
	979	3,082	1,807	5,868

Table 9 shows the positions at workplace for the employees of the three different groups. Most of them are regular employees, around 75% of them in every category, however amongst H&A there more than 20% that has a job not requiring a degree (note: it is a different question than the ones used in the employment definitions). There are very few in the upper management naturally as our group consists of only 25–30 year olds. Otherwise E&IT&EB graduates are more often in management positions, and they also more likely to answer to this question.

Table 10: Weighted observation numbers and fractions of people in the category of position within groups of fields of study

	H&A	E&IT&EB	Rest of non-H&A	Total
missing	105	195	223	524
%	9.8	5.9	11.4	8.3
Upper management	13	33	27	73
%	1.2	1.0	1.4	1.2
Middle management	56	274	113	443
%	5.3	8.3	5.7	7.0
Lower management	58	292	101	451
%	5.4	8.9	5.1	7.1
Employee at a job requiring degree	621	2137	1255	4013
%	57.8	65.0	64.0	63.5
Employee at a job not requiring degree	220	357	243	820
%	20.5	10.9	12.4	13.0
Total	1,074	3,288	1,961	6,323
%	100.0	100.0	100.0	100.0

As for average net salaries Table 10 shows the same pattern that we previously noted except for employees in a job without a degree need where net salaries are around the same, 150 thousand HUFs for everyone. However in every other category H&A graduates earn less, and for the regular employee category their salaries are basically the same regardless of degree requirement. This can explain why H&A have a greater fraction in unqualified jobs: for them based on monthly salaries their degrees have very little return on average so when searching for jobs their preference for qualified job offers are probably lower than for the other categories, so they more often settle for jobs that does not require a degree.

Table 11: Weighted observation numbers and the averages of monthly salaries from primary job by groups of fields of study and categories of position, HUF

	H&A	E&IT&EB	Rest of non-H&A	Total
missing	153,044	223,286	175,046	189,913
N	78	167	182	426
Upper management	208,896	265,289	209,278	235,134
N	12	30	23	66
Middle management	218,114	267,489	226,618	250,631
N	52	252	107	411
Lower management	192,987	302,802	215,238	270,138
N	52	280	93	425
Employee at a job requiring degree	150,197	229,356	166,236	197,433
N	585	2,022	1,183	3,791
Employee at a job not requiring degree	148,964	155,422	142,044	149,788
N	200	330	219	749
Total	156,805	231,246	170,820	200,218
N	979	3,082	1,807	5,868

In the following chapters we are going to look at employment and net salary regressions to estimate the partial correlations with the degree groups controlling for various other factors. For the regressions the dataset provides a huge number of variables of which we included the ones that should have an impact either on the productivity of the person or on his/her labor supply. For missing values we always use a signaling dummy variable. The grouping of control variables for salary regressions are displayed in Table 11, we also give the values these variables can take. In the case of employment regressions from group "work" and "university" variables for previous work experience and type and county of residence municipality are included as naturally there are position etc. variables are not relevant.

Table 12: Groups and values of the regression control variables

socioeconomic	workplace	FEOR	ability proxies	education	university
age (25-30cont.)			type of secondary school (1-4 cat.)	had a degree before graduation (0/1 cat.)	institutions
age squared	type of employment (1-4 cat.)	FEOR groups	type of financing studies (1-3 cat.)	level of previous degree (1-9 cat.)	usefulness of studies at workplace (1-5 cat.)
in relationship (0/1 cat.)	type of contract (1-3 cat.)		grades categories (1-4 cat., 4 is best)	has a degree after graduation (0/1 cat.)	
has child (0/1 cat.)	workplace public/private (0-2 cat.)		grades relative to others in field (1-5 cat., 5 is best)	level of degree after graduation (1-9 cat.)	
grew up abroad (0/1 cat.)	workplace firm size (1-6 cat.)		English language (0/1 cat.)	still studies (0/1 cat.)	
type of municipality at age 14 (1-4 cat.)	employed abroad (0/1 cat.)		German language (0/1 cat.)	level of study (1-9 cat.)	
county at age 14 (1-20 cat.)	type of municipality of workplace (1-4 cat.)		French language (0/1 cat.)	level of degree in question (1-9 cat.)	
highest level of parents' education (1-4 cat.)	county of workplace (1-20 cat.)		Spanish language (0/1 cat.)	schedule of degree in question (1-4 cat.)	
perceived family wealth (1-5 cat.)	category of working hours (1-4 cat.)		Russian language (0/1 cat.)	year of graduation is 2010 vs. 2008	
has family member from study field (0/1 cat.)	employment abroad before graduating (0/1 cat.)		late graduation due to lack of language (0/1 cat.)		
	employment abroad after graduating (0/1 cat.)		number of semesters (cont.)		
	had internship (0/1 cat.)		studied abroad before graduation (0/1 cat.)		
	employed at place of internship(0/1 cat.)		studied abroad after graduation (0/1 dummies)		
	employed in profession before graduating (0/1 cat.)				
	employed not in profession before graduating (0/1 cat.)				
	has secondary job (0/1 cat.)				

Regression analysis results for employment differences

In this chapter we are going to look at the decomposition of the employment differences between humanities and arts students (H&A) and the two other groups of study fields. We are not going to present every coefficient that were included in the regressions but we signal which set of control variables were included. In every case we are going to run separate regressions by gender. We use Huber–White heteroscedasticity robust standard errors provided by Stata's robust option, otherwise we use linear probability models, OLS-regressions. Regression results with the full variable sets can be found in the supplementary material.

First we examine overall employment amongst men using liner probability models. The regression results can be seen in Table 12. The around 4.4 pp difference of H&A against the E&IT&EB group and the 4.0 pp against the rest could be decreased to a non-significant ceteris paribus 3.0 pp and 2.3 pp meaning that the control variables are responsible for around 1/3rd of the difference. If we look at the decrease in the coefficient it seems that the most important set of controls were the group of previous work experience and the group of educational background. Altogether it seems that with regards to overall employment there is no statistically significant difference at p=0.05 level between male H&A and non-H&A graduates. The explanatory power of the final model is still limited with a low R-squared value.

Table 13:Linear probability model results for overall employment amongst male graduates, reference group: H&A

LPM. Dependent variable: Employed. Mean: 95.6. Sample: male graduates, 25–30 years old, employed or unemployed, graduated in 2008/2010.						
E&IT&EB	0.0437**	0.0408**	0.0393**	0.0301	0.0262	0.0302
	(0.0191)	(0.0190)	(0.0189)	(0.0188)	(0.0190)	(0.0207)
Rest of non-H&A	0.0395**	0.0340*	0.0336*	0.0335*	0.0221	0.0232
	(0.0201)	(0.0199)	(0.0195)	(0.0194)	(0.0203)	(0.0219)
Observations	2,642	2,642	2,642	2,642	2,641	2,641
R-squared	0.004	0.042	0.063	0.085	0.108	0.130
included controls:						
socioeconomic		x	x	x	x	x
ability			x	x	x	x
work (pre-employment)				x	x	x
education					x	x
university (only institutions)						x

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Now let's turn to the female graduates as they are the dominant part of the H&A graduate group, the results are displayed in Table 13. Similarly in the beginning there is little raw difference between the groups, controlling for various factors the gap vs. the rest of non-H&A to 3.7 pp, however it increased to 5.5 pp ceteris paribus against the E&IT&EB group. Observing the educational controls the negative partial correlation suggests that altogether relative to their level of education H&A graduates perform even worse, and also including their universities increases the gap. Amongst the control sets the previous work experience seems to decrease the difference the most. The R-squared value for the regression suggests that the model's explanatory power is moderate.

Table 14: Linear probability model results for overall employment amongst female graduates, reference group: H&A

LPM. Dependent variable: Employed. Mean: 94.1. Sample: female graduates, 25–30 years old, employed or unemployed, graduated in 2008/2010.						
E&IT&EB	0.0501***	0.0474***	0.0510***	0.0434***	0.0434***	0.0549***
	(0.0114)	(0.0111)	(0.0113)	(0.0112)	(0.0116)	(0.0134)
Rest of non-H&A	0.0423***	0.0388***	0.0381***	0.0357***	0.0306**	0.0365***
	(0.0119)	(0.0117)	(0.0117)	(0.0119)	(0.0131)	(0.0138)
Observations	4,028	4,028	4,028	4,028	4,028	4,028
R-squared	0.007	0.036	0.052	0.074	0.094	0.103
included controls:						
socioeconomic		x	x	x	x	x
ability			x	x	x	x
work (pre-employment)				x	x	x
education					x	x
university (only institutions)						x

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 14 looks at the differences in employment requiring degree judged by the graduate, amongst males. Even though the model's explanatory power increases with controls the difference stays the same: around 16 pp vs. the E&IT&EB group and an around 8.5 pp against the rest. Including the universities increases the estimated gap between H&A and the two other groups. For this dependent variable the model also seem to have a better explanatory power according to the R-squared.

Table 15: Linear probability model results for employment requiring a degree by the self-judgment definition, amongst male graduates, reference group: H&A

LPM. Dependent variable: Employed in jobs requiring degree (self-judged). Mean: 86.7. Sample: male graduates, 25–30 years old, employed or unemployed, graduated in 2008/2010.						
E&IT&EB	0.161***	0.162***	0.159***	0.154***	0.161***	0.164***
	(0.0301)	(0.0300)	(0.0294)	(0.0291)	(0.0295)	(0.0317)
Rest of non-H&A	0.0879***	0.0879***	0.0791**	0.0819**	0.0684**	0.0840**
	(0.0340)	(0.0335)	(0.0326)	(0.0320)	(0.0343)	(0.0353)
Observations	2,506	2,506	2,506	2,506	2,506	2,506
R-squared	0.024	0.057	0.116	0.153	0.178	0.195
included controls:						
socioeconomic		x	x	x	x	x
ability			x	x	x	x
work (pre-employment)				x	x	x
education					x	x
university (only institutions)						x

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

For females the previous relationships from the overall employment variable repeat themselves as the gap increases with controls compared to the E&IT&EB group from 8.7 to 10.3 pp and decreases compared to the rest from 10.3 to 6.4, as Table 15 reports it. The gap seem to decrease vs. the E&IT&EB group due to socioeconomics background and pre-employment work experience and to increase due to ability, education and universities.

Table 16: Linear probability model results for employment requiring a degree by the self-judgment definition, amongst female graduates, reference group: H&A

LPM. Dependent variable: Employed in jobs requiring degree (self). Mean: 81.1. Sample: female graduates, 25–30 years old, employed or unemployed, graduated in 2008/2010.						
E&IT&EB	0.0868***	0.0810***	0.101***	0.0842***	0.0897***	0.103***
	(0.0183)	(0.0184)	(0.0186)	(0.0183)	(0.0187)	(0.0214)
Rest of non-H&A	0.103***	0.103***	0.0968***	0.0885***	0.0793***	0.0636***
	(0.0186)	(0.0184)	(0.0183)	(0.0179)	(0.0196)	(0.0205)
Observations	3,737	3,737	3,737	3,737	3,737	3,737
R-squared	0.010	0.042	0.092	0.143	0.176	0.191
included controls:						
socioeconomic		x	x	x	x	x
ability			x	x	x	x
work (pre-employment)				x	x	x
education					x	x
university (only institutions)						x

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Our second definition for employment at a job requiring a degree was based on FEOR main groups and included managers, regular employees who needed degrees and also technicians and assistants. Table 16 shows that for males the employment difference stays around 11 pp against the E&IT&EB group (after getting to around 9 without the universities) while it becomes an almost perfect 0 against the rest of the non-H&A graduates. Previous work experience and educational background variables decreased while socioeconomic factors and universities increased the estimated gap.

Table 17: Linear probability model results for employment requiring a degree by the inclusive FEOR-based definition, amongst male graduates, reference group: H&A

LPM. Dependent variable: Employed in jobs requiring degree (FEOR main group is 1 or 2 or 3, so including assistants & technicians). Mean: 92.9. Sample: male graduates, 25–30 years old, employed or unemployed, graduated in 2008/2010.						
E&IT&EB	0.114***	0.116***	0.108***	0.0975***	0.0927***	0.117***
	(0.0244)	(0.0244)	(0.0249)	(0.0245)	(0.0251)	(0.0266)
Rest of non-H&A	0.0122	0.0179	0.00918	0.00869	-0.0197	0.000140
	(0.0296)	(0.0288)	(0.0286)	(0.0280)	(0.0310)	(0.0309)
Observations	2,295	2,295	2,295	2,295	2,295	2,295
R-squared	0.039	0.082	0.127	0.165	0.190	0.223
included controls:						
socioeconomic		x	x	x	x	x
ability			x	x	x	x
work (pre-employment)				x	x	x
education					x	x
university (only institutions)						x

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

For women the situation looks a bit similar, Table 17 shows the regression results. Against the rest of non-H&A group the starting 5.5 pp disadvantage of H&A are decreased to a non-significant 2.8 vs. the rest of non-H&A, while compared to the E&IT&EB it increases from 3.1 pp to 4.5: the inclusion of universities actually increases the gap vs. the latter and decreases vs. the former. The biggest impact except for universities are of the variables of ability group as they increase an only p<0.1 significance 3 pp to a significant 4.5 pp. However the same set of variables seem to explain the variance of the dependent variable significantly worse than in the case of men.

Table 18: Linear probability model results for employment requiring a degree by the inclusive FEOR-based definition, amongst female graduates, reference group: H&A

LPM. Dependent variable: Employed in jobs requiring degree (FEOR main group is 1 or 2 or 3, so including assistants & technicians). Mean: 87.8. Sample: female graduates, 25–30 years old, employed or unemployed, graduated in 2008/2010.						
E&IT&EB	0.0310**	0.0300*	0.0445***	0.0345**	0.0339**	0.0453**
	(0.0155)	(0.0155)	(0.0157)	(0.0156)	(0.0160)	(0.0190)
Rest of non-H&A	0.0553***	0.0526***	0.0485***	0.0460***	0.0374**	0.0280
	(0.0154)	(0.0153)	(0.0154)	(0.0154)	(0.0166)	(0.0171)
Observations	3,552	3,552	3,552	3,552	3,552	3,552
R-squared	0.004	0.031	0.063	0.101	0.121	0.139
included controls:						
socioeconomic		x	x	x	x	x
ability			x	x	x	x
work (pre-employment)				x	x	x
education					x	x
university (only institutions)						x

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

For the third employment definition we only considered those employed who work as managers or employed at a job requiring a degree, but not as assistants or technicians, etc. For this strictest definition Table 18 displays the results amongst males. The around H&A 16 pp difference with the E&IT&EB decreased to 11.7 pp (9.7 pp without universities) and the around 10 pp against the rest is reduced to around 7.0pp (or 2.7pp without universities). Here universities again increase the gap meaning that compared to their fellow graduates of the given university H&A graduates perform worse. Otherwise in this category all of the included other factors decrease the gap between H&A and the others.

Table 19: Linear probability model results for employment requiring a degree by the strict FEOR-based definition, amongst male graduates, reference group: H&A

LPM. Dependent variable: Employed in jobs requiring degree (FEOR main group is 1 or 2, *not* including assistants & technicians). Mean: 74.0. Sample: male graduates, 25–30 years old, employed or unemployed, graduated in 2008/2010.						
E&IT&EB	0.159***	0.154***	0.126***	0.101***	0.0970***	0.117***
	(0.0350)	(0.0347)	(0.0357)	(0.0360)	(0.0373)	(0.0408)
Rest of non-H&A	0.102***	0.105***	0.0814**	0.0702*	0.0266	0.0699
	(0.0394)	(0.0387)	(0.0388)	(0.0388)	(0.0421)	(0.0436)
Observations	2,295	2,295	2,295	2,295	2,295	2,295
R-squared	0.012	0.052	0.115	0.151	0.187	0.215
included controls:						
socioeconomic		x	x	x	x	x
ability			x	x	x	x
work (pre-employment)				x	x	x
education					x	x
university (only institutions)						x

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

For women the strictest definition of employment requiring a degree behaves entirely differently (Table 19). While the H&A advantage vs. the E&IT&EB group diminishes entirely, compared to the rest of non-H&A group it remains 10.2 pp: altogether the 40% of the gap is reduced by the control variables. So looking at the strictest qualified employment definition (which were also used by Köllő, 2015) we get the results for females that compared to the E&IT&EB group H&A graduates have no employment deficit, but vs. the rest a significant disadvantage stays with controls, too, so the exact opposite that we experienced before.

Table 20: Linear probability model results for employment requiring a degree by the strict FEOR-based definition, amongst female graduates, reference group: H&A

LPM. Dependent variable: Employed in jobs requiring degree (FEOR main group is 1 or 2, so *not* including assistants & technicians). Mean: 59.1. Sample: female graduates, 25–30 years old, employed or unemployed, graduated in 2008/2010.						
E&IT&EB	-0.0448**	-0.0463**	-0.0188	-0.0318	-0.0370	0.0101
	(0.0221)	(0.0221)	(0.0229)	(0.0230)	(0.0235)	(0.0282)
Rest of non-H&A	0.170***	0.165***	0.160***	0.159***	0.117***	0.102***
	(0.0220)	(0.0219)	(0.0220)	(0.0223)	(0.0243)	(0.0251)
Observations	3,552	3,552	3,552	3,552	3,552	3,552
R-squared	0.039	0.070	0.112	0.139	0.170	0.198
included controls:						
socioeconomic		x	x	x	x	x
ability			x	x	x	x
work (pre-employment)				x	x	x
education					x	x
university (only institutions)						x

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Overall our results suggest that regarding overall employability there is no significant disadvantage for male H&A graduates and there is a small but statistically significant 3–5 pp disadvantage for female H&A graduates. Regarding employment at jobs requiring a degree the situation depends on the definition. With three different definitions we get amongst males an around 12–16 pp difference vs. the E&IT&EB graduates and a non-significant to 7–8 pp difference with the rest. This suggests that the choice of employment definitions influence the outcomes greatly, however there seems to be a significant disadvantage for H&A graduates against the E&IT&EB group. For females the results are even less clean: the general employment gap for H&A graduates is a small significant 3–5 pp, the self-judged controlled difference in jobs needing a degree is twice this big. But for qualified jobs based on FEOR we received from non-significant to 10 pp deficit for H&A graduates against either other group. It is probable that there is some disadvantage for female H&A graduates however it seems that it depends greatly upon where we put assistants and technicians or whether they can judge their situation for themselves or not, so again definitions are crucial. It is also important to note that neither definition could support the public bashing of H&A graduates, but shows that engineers, people with IT degrees and economics/business graduates seem to perform significantly better in the Hungarian labor market.

Regression analysis results for salary differences

In this chapter we take a look at the monthly net salary differences of the humanities and arts graduates against their two groups of peer graduates. First we examine the salary regression results on the sample with observable salary data. Our dependent variable is the natural logarithm of the monthly net salaries reported by the respondent and we use the whole set of the previously listed control variables. We use standard OLS-regressions with built-in Hubert–White standard errors in STATA. Due to lack of place we only report the coefficients of the field of study groupings, observation numbers and R–squared values and do not show every permutation for the set of controls. We signal the inclusion of a set of variables with a little x.

Table 20 shows how the raw salary difference of around 56% ($e^{0.447} - 1$) against the E&IT&EB and the around 15% salary disadvantage against the rest of non-H&A diminishes to 15% and a non-significant 1.5% respectively. The biggest jump in coefficients happen at the inclusion of work-related controls as more the half of the original gap vs. the E&IT&EB group disappears. If we look at the individual coefficients we can see that position at workplace is a very strong factor in salaries ceteris paribus alongside with foreign ownership and firm size. People at a manager position in big, foreign company earn more money. Altogether we lose around 75% and 90% of the original gaps respectively showing that most of the gap comes from the populations' different base characteristics. The model has a quite good explanatory power at 0.71 R-squared.

Table 21: OLS regression results for log(monthly salary) amongst male graudates, reference group: H&A

OLS. Dependent variable: log(salary). Sample: male graduates, 25–30 years old, employed or unemployed, graduated in 2008/2010.						
E&IT&EB	0.447***	0.439***	0.411***	0.196***	0.179***	0.138***
	(0.0393)	(0.0400)	(0.0383)	(0.0281)	(0.0284)	(0.0346)
Rest of non-H&A	0.137***	0.134***	0.111***	0.0934***	0.0500	0.0150
	(0.0452)	(0.0450)	(0.0424)	(0.0303)	(0.0327)	(0.0396)
Observations	2,174	2,174	2,174	2,174	2,174	2,174
R-squared	0.089	0.169	0.280	0.630	0.646	0.707
included controls:						
socioeconomic		x	x	x	x	x
ability			x	x	x	x
work				x	x	x
education					x	x
university						x
FEORXcountyXcategory of settlement fixed effects						x

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

For females the situation is similar: there is around 13% salary disadvantage for H&A graduates against the E&IT&EB graduates against an around 33% raw difference and non-significant 1.4% deficit vs. the rest with controls on (Table 21). The situation is also quite similar that the gap decreases the most with the work variable group against the E&IT&EB, however it increases the gap vs. the rest of the non-H&A group. The final model's R-squared value of 0.725 suggests that our model has a good explanatory power.

Table 22: OLS regression results for log(monthly salary) amongst female graudates, reference group: H&A

OLS. Dependent variable: log(salary). Sample: female graduates, 25–30 years old, employed or unemployed, graduated in 2008/2010.						
E&IT&EB	0.286***	0.268***	0.293***	0.151***	0.153***	0.122***
	(0.0241)	(0.0239)	(0.0232)	(0.0164)	(0.0168)	(0.0207)
Rest of non-H&A	0.0206	0.0248	0.0316	0.0543***	0.0222	0.0142
	(0.0246)	(0.0241)	(0.0224)	(0.0152)	(0.0160)	(0.0182)
Observations	3,202	3,202	3,202	3,202	3,202	3,202
R-squared	0.065	0.159	0.291	0.670	0.682	0.725
included controls:						
socioeconomic		x	x	x	x	x
ability			x	x	x	x
work				x	x	x
education					x	x
university						x
FEORXcountyXcategory of settlement fixed effects						x

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

If we look at only those who work in Hungary the estimated controlled differences for males increase suggesting that working abroad helps H&A graduates to receive closer salaries to their peers (Table 22). Against E&IT&EB from an around 60% ($e^{0.473} - 1$) disadvantage for H&A graduates we arrive to an around 20% disadvantage with controls on, and for the rest it is an around 17% taw difference that turn into a non-significant 6.5% for males. We again see that the most important group of variables are the ones in the "work" grouping with position, foreign ownership and firm size playing an important role. The R-squared of the model is 0.65, so the explanatory power of the model is good.

Table 23: OLS regression results for log(monthly salary) amongst male graudates working in Hungary, reference group: H&A

OLS. Dependent variable: log(salary). Sample: male graduates, 25–30 years old, employed or unemployed, graduated in 2008/2010 working in Hungary.						
E&IT&EB	0.473***	0.478***	0.454***	0.224***	0.209***	0.183***
	(0.0332)	(0.0330)	(0.0332)	(0.0281)	(0.0287)	(0.0349)
Rest of non-H&A	0.159***	0.166***	0.141***	0.115***	0.0791**	0.0645
	(0.0380)	(0.0372)	(0.0367)	(0.0305)	(0.0327)	(0.0395)
Observations	1,994	1,994	1,994	1,994	1,994	1,994
R-squared	0.125	0.219	0.303	0.558	0.580	0.652
included controls:						
socioeconomic		x	x	x	x	x
ability			x	x	x	x
work				x	x	x
education					x	x
university						x
FEORXcountyXcategory of settlement fixed effects						x

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Amongst females the raw H&A salary disadvantage against E&IT&EB graduates is 37% and non-significant 3.9% against the rest (Table 23). When controlling for socioeconomic background and ability the gap actually widens a bit, work group of variables explain half of the original gap of the bigger difference. In the final configuration 13% disadvantage is left against the E&IT&EB and a non-significant 2% against the rest. The final model has an explanatory power of 0.65 R-squared.

Table 24: OLS regression results for log(monthly salary) amongst female graudates working in Hungary, reference group: H&A

OLS. Dependent variable: log(salary). Sample: female graduates, 25–30 years old, employed or unemployed, graduated in 2008/2010 working in Hungary.						
E&IT&EB	0.312***	0.297***	0.320***	0.152***	0.155***	0.130***
	(0.0209)	(0.0202)	(0.0204)	(0.0165)	(0.0169)	(0.0209)
Rest of non-H&A	0.0393*	0.0481**	0.0436**	0.0490***	0.0218	0.0198
	(0.0203)	(0.0195)	(0.0189)	(0.0149)	(0.0156)	(0.0177)
Observations	2,942	2,942	2,942	2,942	2,942	2,942
R-squared	0.102	0.206	0.294	0.578	0.594	0.655
included controls:						
socioeconomic		x	x	x	x	x
ability			x	x	x	x
work				x	x	x
education					x	x
university						x
FEORXcountyXcategory of settlement fixed effects						x

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

So according to our salary OLS-results we can observe that amongst both men and women humanities and arts graduates have a net salary disadvantage of around 13–20% compared to the group of engineers, IT graduates and economics/business graduates while the deficit vs. the rest of the non–H&A is small and statistically not significant at the 5% level. It reinforces our previous remarks that humanities and arts graduates perform worse only to the former group.

Robustness and discussion of results

In this chapter we are going to look at some other estimations to examine the sensitivity of our results, and afterwards we are going to discuss the potential biases that we left untreated and might influence our results. We already used three different definitions for employment gap and gap in employment requiring a degree, however we can check whether our results look similar using logit regressions' average partial effects instead of linear probability models. We only report the final specifications for our equations in Table 24:

(1): the self-judged definition of degree requirement

(2): the more inclusive FEOR-based of degree requirement (FEOR-groups 1, 2 and 3)

(3): the more restrictive FEOR-based definition of degree requirement (FEOR-groups 1 and 2)

For males it seems that logit APE-s are around the same as for the linear probability models' average treatment effects vs. the E&IT&EB and the rest of non-H&A, however statistical significance of the coefficients changed. This might also be due to not weighting in logit estimations. So compared to the engineering, information technology, economics and business graduates humanities and arts graduates have an around 13–16 pp lagging behind in qualified employment, while compared to the rest it is around 0-9 pp.

Table 25: Average partial effects of logit estimations with full control set on for male graduates by the three different definitions of employment requiring a degree, reference group: H&A

	(1)	(2)	(3)
E&IT&EB	0.152***	0.130	0.160***
	(0.0305)	(0.108)	(0.0370)
Rest of non-H&A	0.0817**	0.0264	0.0901**
	(0.0324)	(0.0352)	(0.0385)
Observations	2,386	1,721	2,176

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

For women the same coefficients for the full control set can be seen in Table 25. Comparing these number to the OLS-results we can see that there are no significant differences regarding the employment gaps, however significance of the variables changed a bit, however they support our previous results.

Table 26: Average partial effects of logit estimations with full control set on for female graduates by the three different definitions of employment requiring a degree, reference group: H&A

	(1)	(2)	(3)
E&IT&EB	0.0864***	0.0411	0.0307
	(0.0192)	(0.0257)	(0.0254)
Rest of non-H&A	0.0612***	0.0219	0.108***
	(0.0183)	(0.0187)	(0.0227)
Observations	3,691	3,441	3,517

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

To check the robustness of our salary estimations we tested our results for different sample restrictions. These were:

(1): Without weighting

(2): Graduates only that work 30–49 hours

(3): Graduates only that work 30–49 hours and work in Hungary

(4): Graduates only that work 30–49 hours, work in Hungary and not teachers

For males Table 26 shows the key coefficients for the final equations with the full control set on. It shows that weighting did not influence our results greatly, however a lot depends on the sample that we choose. With stricter sample restriction the salary disadvantage for H&A graduates seem to increase against both other groups: the more conventional employees we target the bigger the differences grow. However our results stay unchanged: there is an around 15-20% statistically significant salary disadvantage for humanities and arts graduate males against engineer, IT and economics/business graduates and a small, not significant disadvantage vs. the rest.

Table 27: OLS-regression results of the key explanatory variables on the dependent variable log(monthly salaries) with full control set on male graduates of four different subsamples, reference group: H&A

	(1)	(2)	(3)	(4)
E&IT&EB	0.142***	0.164***	0.207***	0.210***
	(0.0349)	(0.0366)	(0.0370)	(0.0373)
Rest of non-H&A	0.0109	0.0323	0.0716*	0.0764*
	(0.0390)	(0.0407)	(0.0402)	(0.0411)
Observations	2,174	1,899	1,756	1,726
R-squared	0.689	0.720	0.668	0.662

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

For women our previous conclusions stand even more firmly: almost regardless of the sample specification the salary disadvantage for H&A stays around 13-15% against the E&IT&EB graduates and around a non-significant 0-2% showing that for women our results appear stable and not dependent on the sample.

Table 28: OLS-regression results of the key explanatory variables on the dependent variable log(monthly salaries) with full control set on female graduates of four different subsamples, reference group: H&A

	(1)	(2)	(3)	(4)
E&IT&EB	0.132***	0.122***	0.137***	0.132***
	(0.0218)	(0.0206)	(0.0222)	(0.0233)
Rest of non-H&A	0.000403	0.0226	0.000679	0.00577
	(0.0188)	(0.0178)	(0.0185)	(0.0210)
Observations	2,770	3,202	2,568	2,343
R-squared	0.742	0.720	0.665	0.658

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

We have to face to following potential biases when we evaluate our results:

- **Response rate bias**: as already mentioned responding is selective and might depend greatly on the added value of the university for the graduate, resulting in the higher response rate for the more successful graduates.
    - Let us consider an example of 100–100 graduates where successful graduates all answer and only 2/3rd of "failed" graduates answer for both groups. If one group of graduates have in reality a higher "failure rate"(for example: 30 people vs. 15 people, 30% vs 15%) with similar non-answering rates (only 20 and 10 people in the sample) the raw gap (15pp) is underestimated (20/(20+70)~22% vs. 10/(10+85)~11%, around 11pp gap). If H&A graduates in fact have a lower "success rate" and our premises stand, we underestimate the gap, however we do not know the actual fractions.
- **Selection bias by ability regarding fields of study**: we could not control for ability properly, we could only use proxies like the type of secondary school, type of financing studies, knowledge of languages or grades. If H&A graduates have on average worse abilities and ability is positively correlated with labor market success then we overestimate the gap. However when we look at the level of English, relative grades compared to their peers or secondary school types (fraction attending six or eight class secondary schools) we find that altogether there is no great difference.
- **Selection bias of having salary data**

If having salary data is not random, then it might result in biased estimation of the coefficients, here salary offers: in our case when the selection is non-random (the person's unemployment is not random or it is systematic when respondents fail to provide an answer), our estimates

will not be unbiased (Kézdi, 2004). The goal of the Heckman selection model is to correct for the probabilities of being selected, we use STATA's "heckman" command with the two-step procedure for estimation. The selection equations include the same variables as for employment with the change that having a partner variable is present only in the selection equation as it is assumed to influence only employment directly (e.g. through connection networks) but not directly salary offers (as it does not influence productivity). Heckman selection model's results are in Table 28. It shows that for males there is a significant around 20% advantage for E&IT&EB graduates vs. humanities and arts graduates whereas there is not any significant advantage for the rest of the non-H&A group[9]. This result suggests that the model excluding those who work outside of Hungary is less biased. For women the H&A disadvantage vs. E&IT&EB graduates is around the 10% (it seemed to be a bit positively biased) and against the rest still non-significant and small.

[9] We were not able to run the model for males separately due to computational limitations, so we report the coefficients from the equations including women and controlling for them with interaction terms.

Table 29: Heckman selection model's result for the dependent variable log(monthly salaries), full set of controls, reference group: H&A (remarks: male estimations are calculated using both genders and interaction terms)

	Male	Female
E&IT&EB	0.184***	0.105***
	(0.0260)	(0.0207)
Rest of non-H&A	0.0254	0.0155
	(0.0277)	(0.0174)
lambda	0.0380	-0.0981
	(0.0668)	(0.0768)
Observations	6,669	4,028

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Conclusions

The paper examined the employment, employment requiring tertiary education and primary monthly salary differences between humanities and arts (H&A) graduates and the others using the 2013 database of the Hungarian Graduate Career Tracking System (GCTS, Nyüsti–Veroszta, 2014) provided to MKIK GVI by the Educatio Public Services Non-profit LLC. The data have several limitations originating from uneven response rates amongst universities and other factors, we intended to make corrections regarding the quality of the dataset to focus more on the target group of young graduates and to lessen systematic bias. We separated the non-H&A into two groups: engineering, information technology and economics/business graduates (E&IT&EB) and the rest. We looked at the raw differences of the different groups by descriptive aspects like genders, working hours, positions at workplace where it seemed that H&A graduates have a raw salary disadvantage versus each group and they also seem to have jobs more often that do not require a degree.

We estimated employment and wage differences using four different employment definition and for many different subsamples to check for the sensitivity of our results. Regarding employment our results show that there are no significant conditional overall employment differences across groups for males, and there is only an around 5-6 pp gap for females vs. the E&IT&EB group and a 3-4 pp gap vs. the rest. As for having a job requiring tertiary education H&A male graduates have an around 12–16 pp disadvantage against E&IT&EB graduates, and an around 0–7 pp against the rest, depending heavily on the definition. For women definitions matter even more: results range from 0–10 pp vs. E&IT&EB graduates and also 0–10 pp vs. the rest, so for women employment differences are rather unclear. Regarding conditional salary differences we found that ceteris paribus H&A graduates have a disadvantage only vs. E&IT&EB graduates of around 13-20% for males and an around 10-13% for females. The differences to the rest of non-H&A graduates appear to be non-significant in either cases. The results seem to vary for males depending on samples, however for women they seem the same, however selection plays a greater role in case of females.

Overall our results suggest that common beliefs, which are widespread in Hungary about the terrible labor market conditions for humanities and arts graduates in Hungary do not hold very well. Most of the gap compared to other graduates are due to the better than average labor market situation for engineering, information technology and economist/business graduates and even compared to them the scale of disadvantage is not enormous. The real story seems to be more about the high returns for these graduate groups rather than the miseries of the humanities and arts graduates.

References

Ábrahám, Árpád and Gábor Kézdi (2000): Long-run trends in earnings and employment in Hungary, 1972–1996. *Budapest Working Papers on the Labour Market*. BWP. 2000/2

Berde, Éva (2010): Óvatosan a diplomás pályakövetés elvárásaival! *EDUCATIO* 2010/3. 448–459.

Garai, Orsolya (2014): *DPR Monitoring 2014.* Educatio Társadalmi Szolgáltató Nonprofit Kft.

Hajdu, Tamás, Zoltán Hermann, Dániel Horn, Gábor Kertesi, Gábor Kézdi, János Köllő and Júlia Varga (2015): Az érettségi védelmében. *Budapest Working Papers On The Labour Market*. 2015/1.

Kézdi, Gábor (2004): Az aktív foglalkoztatáspolitikai programok hatásvizsgálatának módszertani kérdései. *Budapest Working Papers On The Labour Market*.2004/2.

Köllő, János (2015): Nincs is túltermelés bölcsészekből. *Defacto Blog* Entry February 16th, 2015. URL: http://index.hu/gazdasag/defacto/2015/02/16/nincs_is_tultermeles_bolcseszekbol/ Access: August 4th, 2015.

Micklewright, John, Sylke V. Schnepf and Chris Skinner (2010): Non-Response Biases in Surveys of School Children: The Case of the English PISA Samples. *IZA Discussion Paper* No. 4789

Nyüsti, Szilva and Zsuzsanna Veroszta (2014): *Hungarian Graduate Career Tracking 2013 Integration of Administrative Databases*. Educatio Public Services Non-profit LLC, Department of Higher Education.

Penwarden, Rick (2014): *Response Rate Statistics for Online Surveys -What Numbers Should You be Aiming For?* URL: http://fluidsurveys.com/university/response-rate-statistics-online-surveys-aiming/ Accessed August 2nd, 2015.

SurveyGizmo (2015): Survey response rates (Jul 27, 2015). URL: https://www.surveygizmo.com/survey-blog/survey-response-rates/ Accessed August 2nd, 2015.

Varga, Júlia (2010): A pályakezdő diplomások keresete, munkaerő-piaci sikeressége a 2000-es évek elején. In: Garai O.–Horváth T.–Kiss L.–Szép L.–Veroszta Zs. (szerk): *Diplomás Pályakövetés IV. Frissdiplomások 2010*. Educatio Társadalmi Szolgáltató Nonprofit Kft. Felsőoktatási Osztály, 59–81.

Varga, Júlia (2011): A pályakezdő diplomások munkaerő-piaci sikeressége 2011-ben. In: Garai O.– Veroszta Zs. (szerk): *Frissdiplomások 2010*. Educatio Társadalmi Szolgáltató Nonprofit Kft. Felsőoktatási Osztály, 143-167.

Veroszta, Zsuzsanna (2014): *Diplomás pályakövetési adatok 2013. Intézményi adatfelvételek.* Educatio Társadalmi Szolgáltató Nonprofit Kft.