Top Earnings Inequality and the Gender Wage Gap in Hungary

by Edina Gábriel

Submitted to

Central European University

Department of Economics and Business

In partial fulfillment of the requirements of the degree Economics MA

Supervisor: Professor László Mátyás

Budapest, Hungary

2018

Abstract

This thesis analyzes the connection between the gender inequality among top earners and the overall gender pay gap in Hungary. I perform this analysis by dividing the earnings distribution into four mutually exclusive centile groupings in the following way: the bottom 90%, next 9%, next 0.9% and top 0.1%. Using these centile groupings I first analyze the female shares and actual annual earnings in each part of the distribution. Then, I build counterfactuals that control for the differences in gender shares in each of the four centile groupings, resulting in a simulated wage distribution that assumes the same share of women that the actual share of men is in each of the centile groupings. Then I use the Oaxaca-Blinder decomposition to learn how much part of the gender wage gap the centile groupings account for. Doing this, I find that the counterfactual earnings ratio – unlike in other developed countries – gives less favorable results for women in Hungary. Furthermore, it turns out that using the widest possible model, the inequality in Hungary is higher than the difference in earnings, thus having the centile groupings to explain more of the gap than the gap itself.

Acknowledgements

I would like to express my gratitude to my supervisor, Professor László Mátyás, whose useful recommendations were of great help to me during writing my thesis. Furthermore, I am thankful to Professor Gábor Kőrösi for helping me to get the right type of data for my analysis and also giving me useful advice about the methodological approach and valuable insights about the data source. Finally, I am beholden to Thomas Rooney from the Center for Academic Writing who gave me substantial pieces of advice regarding the structure and style of my thesis.

Contents

List of Figures

List of Tables

Chapter 1 – Introduction

In the past 20-30 years, the share of women in the labor force has gone up in most developed countries. However, the income inequality between men and women is still significantly large. The increasing share of women in the labor force does not necessarily mean that their share is the same in every occupation or in every decile or percentile of the earnings distribution, which – of course – cannot really be a sensible expectation. Traditionally emerged shares still seem to hold in most developed countries, so that the share of women in the lower part of the distribution is still higher. There are sticky floors and glass ceiling effects that still keep women from having the same chances in the labor market as men.

One measure of the inequality between the income of women and men is the gender wage gap. Another way of measuring inequality is to analyze the share of women in the labor force, and more specifically, in different parts of the earnings distribution. This is a valid method, because the share of women tend to be very different in the different centiles of the distribution. For example, it is usually higher in the bottom deciles, while in the top percentiles it tends to be far lower even in developed countries.

To illustrate the importance of the above mentioned trends in the share of women it is worth mentioning that currently some leading international companies sometimes prefer hiring women in some of their teams to get a more balanced gender composition. Specifically, an interesting question raises about how is the share of women among top earners (in the top 0.1% of the earnings distribution) related to the overall gender wage gap.

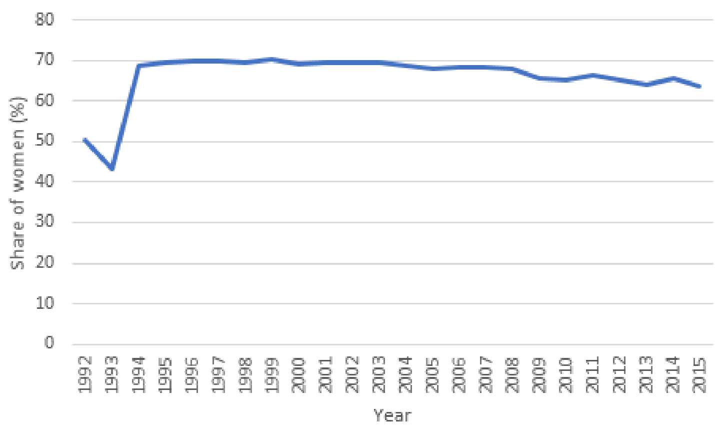
In several developed countries there were substantial increases in the share of women among top earners in the past decades. For example, Guvenen et al (2014) documented that the share of women in the top 0.1% of the wage distribution increased from 1.9% to

10.5% between 1985 and 2012. Atkinson et al (2016) found similar tendencies of the share of women in top income groups using data on Australia, Canada, Denmark, Italy, New Zeland, Norway, Spain and the United Kingdom.

Fortin et al (2017) are the first to examine the relation between the share of women among top earners and the overall gender pay gap. They use data on Canada, Sweden and the United Kingdom. In my thesis I do a similar analysis on Hungarian data spanning the years 1992-2015, which has not been done so far for this country. Throughout the analysis I follow their steps in terms of descriptive statistics as well as in empirical methodology. One difference between the two analyses is that while they perform some robustness checks for the centile groupings they divide the wage distribution into, I use only the one they detail in their paper and that has spread in the literature (Brewer et al, 2008; Guvenen et al, 2014).

Figures 1.1 and 1.2 show the overall share of women in the Hungarian labor force and the share of women in the top 0.1% of the wage distribution respectively.

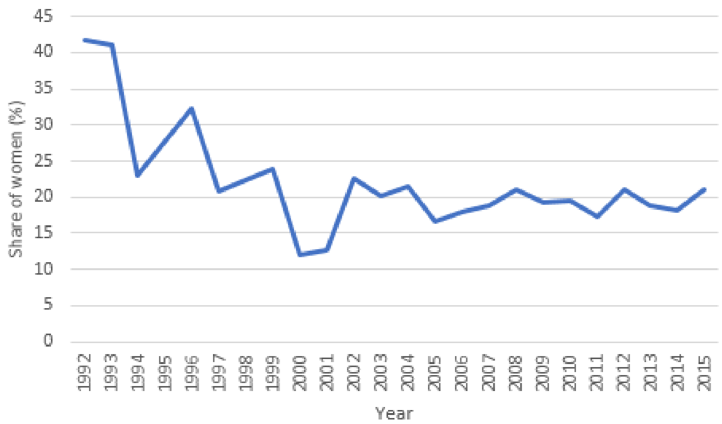
Figure 1.1: The overall share of women in the labor force in Hungary



Source: Own calculations from the Hungarian wage survey data

An interesting feature of the Hungarian data is that the overall share of women has somewhat decreased in the past years unlike in most European countries; it was around 50% in 1992, then after a sudden drop in 1993 it increased to around 70% in the following year, and it stayed at around that level with a slight decrease beginning in 2003-4. Using the total yearly income of employees, the share of women in the top 0.1% of the wage distribution has also decreased in these years. This is a much larger decrease than that

Figure 1.2: The share of women in the top 0.1% of the wage distribution in Hungary



Source: Own calculations from the Hungarian wage survey data

of the overall share starting from 42% in 1992, while ending up at around 20% in 2015. Even though these changes are unlike those found in other developed countries, the above mentioned change is economically large enough to give the basis for the analysis.

Like Fortin et al (2017) I also use counterfactual analysis, namely the Oaxaca-Blinder decomposition, to derive conclusions about the connection between the change in the share of women among top earners and the overall gender pay gap. I implement this by building counterfactuals that assume the share of women in centile groupings of the wage distribution (detailed later) are the same as the share of men, which helps to identifying the part of the wage gap that can be explained by the gender inequalities. Then, I use the usual Oaxaca-Blinder decomposition in two of the available years, using two different models to explain the wage gap. A large change in the share of women in the top 0.1% of the wage distribution should be able to explain a substantial part of the overall gender wage gap. (This hypothesis is based on the results of Fortin et al (2017).)

The remainder of the paper is organized as follows. Section 2 gives an overview on the relevant literature, while Section 3 presents the data used. Section 4 presents the empirical analysis. Section 5 concludes and gives a comparison of results to those of Fortin et al (2017).

Chapter 2 – Literature Review

The literature I mainly use for my analysis is the paper by Fortin et al (2017) who analyze the relationship between the top earnings inequality and the overall gender pay gap in Canada, Sweden and the United Kingdom using administrative data in order to capture the upper tail of the income distribution with annual earnings as a measure of income. They build their analysis on counterfactuals asking "what if women were represented at the top of the earnings distribution of men and women combined the same way men are?". They use the classical Oaxaca-Blinder decomposition method to address their research question and assess the relative explanatory power of different factors on the wage differences. Gender-specific OLS regressions are to give the basis of this empirical methodology.

My analysis follows Fortin et al (2017) in most aspects. I have a data set that allows me to perform the same empirics, while I have the same problem in terms of the measurement of earnings like they have in case of Sweden. Even if both data contains the actual level of earnings, their big disadvantage is the lack of available working hours that makes the results somewhat harder to interpret. While they perform a couple of robustness checks with different divisions of the wage distribution when calculating the counterfactual earnings ratios and find no major differences in results, I rely on the centile groupings they do their main analysis with and I dispense with checking the other, slightly different, divisions of the distribution.

Their main results are the following. For Sweden, they find that educational attainment has a negative effect on earnings, which means the failure of the classical human capital model (in terms of the gender pay gap) and the rejection of the Mincer-Polachek hypothesis (Mincer and Polachek, 1974). According to this hypothesis, the human capital of women has a shorter pay-off period than that of men. Women usually take some time off (e.g. because of childbearing). As they have a shorter pay-off period, they are not as motivated to accumulate human capital as men are. Moreover, their human capital depreciates during the time they are not working. So, in the end, women tend to acquire less human capital than men. However, the results of the paper show that the effect of human capital on earnings is negative, while the payoff-time theory would work only in case of positive effects. Another finding is that the explanatory power of the centile grouping remains large.

Finally, there is a growth in the explanatory power of the centile groupings over time. As the centile groupings division is used to measure the effect of inequality in given parts of the earnings distribution, it turns out that inequality is becoming more important over the years in explaining the pay gap. This can also be expressed in terms of "swimming upstream", a term introduced by Blau and Kahn (1997). [1] These main findings are also applicable to the other two countries they studied.

Based on these findings they also highlight some policy implications and effectiveness of existing public policies. I will return to their main empirical results in sections 3 and 4 giving a comparison with those found using data on Hungary.

Several papers discuss the general trends in the relevant data worldwide. Pikkety and Saez (2003) document the increase in top income share in the United States, for example the income share of the top 0.1% increased from 2.02% in 1976 to 6.04% in 1998. In Pikkety and Saez (2013) we can find information on more recent years. They document inequality exceeding early 20th century levels given that the inequality accruing to the top 10% reached 50% in 2012. Alvaredo et al (2013) find that the income share of the top 1% has increased from 9% in 1976 to 20% in 2011 in the US and similar numbers in Canada and the United Kingdom.

Davies (2015) documents that in the United Kingdom the share of women in top positions was only around 25% in 2015 despite the increasing share of women in labor force, thus motivating an analysis of the effects of the inequality in the top tail of the wage distribution. Guvenen et al (2014) also document increasing top earning shares as well as

[1] "Swimming upstream" refers to the intention of women to pay equality despite the factors that makes this difficult and slows the progress of narrowing the gender wage gap.

show the increasing share of women in top income groups. They find that the share of women in the top 0.1% increased from around 5% in 1980 to above 10% in 2010. Atkinson et al (2016) show similar trends in the share of women for eight advanced economies.

In terms of increasing inequality, Blau and Kahn (1997) have written about the struggling of women for equality, marking it with the term 'swimming upstream' as showing the difficulty of reacing it especially that residual inequality has increased as well.

For recent increases in top incomes, Pikkety et al (2014) and Gabaix and Landier (2008) came up with explanations involving tax rates and global economic considerations.

In terms of methodology I use results from Brewer et al (2008) and Guvenen et al (2014). They use the method of dividing the distribution into mutually exclusive centile groupings. Brewer et al (2008) use the following centile groupings: top 10-1%, top 1-0.1% and top 0.1%, and compare these to the group of all taxpayers. Guvenen et al (2014) use the lower 99%, next 0.9% and top 0.1% centile groupings.

One reason behind my analysis is that Fortin and Lumieux (2000) show that women tend to be present in a higher share in the lower tail of the wage distribution and in a lower share in the upper tail. This supports the idea of women having difficulties reaching the top positions, however this does not claim anything about the increasing trend of women in the upper tail in most countries.

Following Fortin et al (2017) I also depart from the traditional methods used in papers like Albrecht et al (2003) and Arulampalam et al (2007) who use earnings differences as their methodology and belong to the classical glass ceiling effects literature.

Edlund and Kopczuk (2009) and Gneezy and Pietrasz (2013) provide explanations for the decreasing gender earnings ratio in the top 0.1% of teh wage distribution. they claim that it might be due to the decline in inherited wealth or the oil and gas extraction as a source of enrichment, and the women with lower abilities being drawn into the top 0.1% because of the lower supply of women with the required abilities, respectively.

I will build counterfactuals like the one of Mincer and Polachek (1974) who asked "What if women had the same level of human capital (e.g. educational attainment and labour market experience) as men?" (also cited by Fortin et al (2017) the same way). They used this counterfactual to test if inequality decreases when women will have the same amount of human capital as men.

Like Fortin et al (2017), I will also use the following counterfactuals: "what would be the average wage of women if they had the same characteristics as men?", and "what would be the average wage of men if they were paid as women?"

Kline (2011) show that these counterfactuals can be computed as regression coefficients as well as by re-weighting used by DiNardo et al (1995).

Blau and Kahn (2017) suggest that complementary explanations to extensive set of variables like human capital measures, occupation, etc. to the gender wage gap, may include gender differences in the wage distribution (as in Fortin et al (2017)). They also show that the gender wage gap is 62% in the US, while Baker and Drolet (2010) document a 67% gap in Canada in 1997, and only a 55% gap for the US in 2010.

In Hungary, Frey (1997) discusses the post-transition trends in the share of women in the labor force. She explores the reasons behind its huge drop in the early 1990s, both overall and dividing the sample into different categories. Using purely statistical indicators, she claims that although the share of women in the labor force greatly decreased in the years between 1990 and 1997, there were also advantages to this trend. There were improvements in the age composition of those who remained active in the labor force. Moreover, there were also improvements in the education level of these women. This made it possible for them to participate in the socio-economic movements after the transition, and led to an increase in the share of women in top positions between 1990-1993. The wage gap also decreased during these years.

Cukrowska-Torzewska and Lovász (2016) conducts a study about how much children contribute to the gender wage gap. They make a comparison of Poland and Hungary using the Oaxaca-Blinder decomposition with simultaneous corrections for selection into employment and parenthood. However, they do not find very significant effects on the gender wage gap. For Hungary they use the Household Budget Survey data for the analysis.

Lovász (2016) also focuses on mothers' employment and its connection with children, more specifically the expansion of childcare in post-socialist countries. She finds that extended childcare has a positive effect on mothers' employment decisions. On the other hand, she points out that in these countries, cultural norms against the employment of women and financial difficulties can constrain the effectiveness of the otherwise successful policies. In terms of data, she does not use the one collected in Hungary, but the OECD family database which contains all countries she focuses o.

Lovász and Telegdy (2010) use a new method to measure discrimination instead of the traditional individual level wage equations: a measure of productivity on the group level. They find that the productivity of women is not lower than men, but they are underpaid according to the individual wage results.

Lovász (2014) does an analysis on the gender wage gap focusing on occupational segregation in Hungary. In her analysis, she uses individual wage equations including observable worker and employer characteristics, gender dummies and sector indicators. The main finding of the paper is in accordance with her hypothesis: discrimination turns out to be smaller in the public sector.

In chapter 3 I present the data set, give more details about the centile grouping used in section 4 and the trends in earnings and the share of women in different centile groups of the wage distribution.

Chapter 3 – Data

3.1 Data Set

I use administrative data on earnings and details of employment provided by the Hungarian Academy of Sciences, Center for Economic and Regional Studies (MTA, KRTK). This is an matched employer-employee data set collected by the National Employment Office and is sampled the following way. From the budget sector all institutions are taken and sampling happens within institutions. For the private sector, all companies are taken above a certain size and random sampling happens from among the employees of the companies. Under this certain size, random sampling happens at the company-level, while all employees of the sampled companies are taken into the sample. In those cases where sampling happens within the company, the employer should provide every standard employment indicator for employees born on the 5th, 15th or 25th of a month. The data is collected in May each year.

One great advantage of the data being collected this way and not using tax data is that this way the data is not top-coded [1] which is a key feature for this analysis.

Overall, 100-200 thousand employees are in the sample each year, provided by the employers, which means that about 10% of the population is sampled.

The sample includes the years between 1992-2015, and all necessary variables are available for the analysis, for example, income, age, education, experience, etc.

In order to get a representative sample individual weights are provided for all indi-

[1] A variable is top-coded if data points that are above a certain value are censored. For example, top-coding is used in tax data, where above a certain level of earnings, the actual earnings is not available in the data, only that certain upper-bound value. Top-coding is usually used in survey data to preserve the anonimity of respondents, as it in some cases it would not be very difficult to find out who is the person with highly outlying values of given variables.

viduals in the sample in each year. These weights represent the share of that kind of observation, with those certain characteristics and values of the variables, in the population. So, for example, if we want to calculate the average earnings in the population, then these individual weights should be used in order to get the average earnings not only for the individuals who are sampled, but an estimated average earnings value for the population as a whole.

My analysis, however, contains weights that are needed for the re-weighting of the individuals in the different centile groupings to get the counterfactual earnings ratios, which are simulated ratios showing what would be the case if the share of women in each of the centile groupings was the same as that of men. The fact that I have to use weights as part of the analysis means that using the individual weights mentioned in the previous paragraph would mean double-weighting in case of the counterfactuals, which over-complicates the analysis. As the sample is relatively large enough I decided to ignore the individual weights needed for the measures of central tendency, regressions etc. to be representative for the population, without loss of quality of the results or a significantly large difference in them with high probability.

For comparability reasons, as in Fortin et al (2017) I also restrict my analysis to workers between 25-64 years of age, so that I also get a sample of those who are more attached to the labor market, excluding young people who are more likely to work part time while studying and pensioners. Moreover, I also correct for inflation to be able to use the real wage changes that are cleaned from price changes effects. For this I use data on CPI from the Hungarian Central Statistical Office (KSH, `www.ksh.hu`).

Considering difficulties with data, I have the same problem of inappropriate data as Fortin et al (2017). I have administrative data which, even though it is collected from the employers directly, does not contain hours worked. This means we cannot distinguish between discrimination at the workplace and the fact that women usually work less hours than men on average due to housework distribution reasons and reasons related to their children. Using the available measure of the gender pay ratio, we can draw conclusions only about income shortfalls caused by the above mentioned reasons, but it is impossible to derive conclusions about how employers treat women differently from men.

In their analysis, Fortin et al (2017) use data from the years 1990 and 2010 for Sweden and 1997 ad 2015 for Canada ad the United Kingdom when applying the Oaxaca-Blinder decomposition. Due to data availability reasons as well as making an attempt to do as a relevant and recent analysis as possible, I will use the years 2000 and 2015.

3.2 Descriptive Statistics

In this section I discuss some descriptive statistics trends in the data which gives the motivation to the analysis.

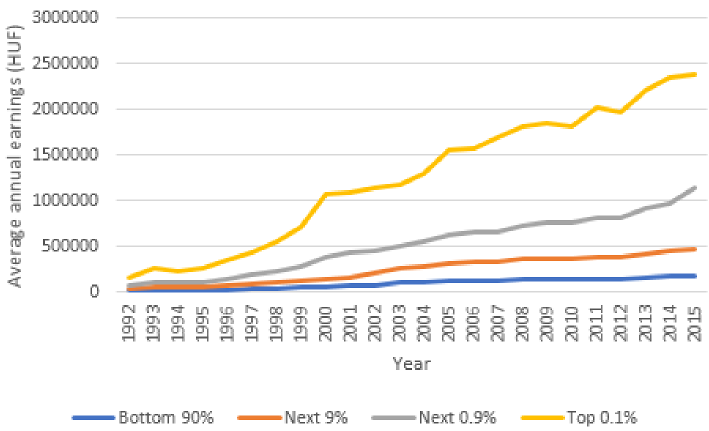
First, based on the methodology of Fortin et al (2017) I also divide the joint distribution of men and women into mutually exclusive centile groupings as seen in Brewer et al (2008) and Guvenen et al (2014), but using four centile groupings as in Fortin et al (2017). Doing this, I have the following centile groupings: bottom 90%, next 9%, next 0.9% and top 0.1%. Doing this makes it possible to perform an analysis that distinguishes between different parts of the earnings distribution, thus allowing me to estimate to what extent inequality in the different centile groupings explain the gender pay gap. For example, Oaxaca decompositions can be performed to see how belonging to a given centile grouping relates to the gender wage gap and how much of it belongs to that part of the earnings distribution.

Using these centile groupings I can build counterfactuals that tell us "what if the share of women in the top 0.1% of the wage distribution was the same as the share of men?" Similar positional ranks were also used by Fortin and Lumieux (1998) and Bayer and Charles (2016).

Figure 3.1 shows the average annual hourly earnings for the four centile groupings I divided the sample into. Earnings are measured in HUF and kept in levels to avoid the shrinking effect of the logarithm on the top.

The figure shows that the annual earnings of the bottom 90% has not increased much relative to the other centile groupings. The earnings of this part of the wage distribution have increased only 11% yearly. The next 9% and 0.9% show some increase, but these increases, although substantially large, still seem to be slight compared to the increasing earnings trend of the top 0.1% of earners. In the next 9% of the distribution annual

Figure 3.1: Average annual earnings trends by centile groupings



Source: Own calculations from the Hungarian wage surwey data

earnings have increased 12% per year on average, while in the next 0.9% the average increase is still 13% per year. Surprisingly, while the trend in the top 0.1% seems to be growing much more dynamically than in the other centiles of the distribution, the average annual increase in earnings is the same in this group as in the previous one: 13% yearly.

These changes are mostly as one would expect. Even if the change in the bottom 90% of the wage distribution has been large during the years, it is still far below the wage of top earners. However, it is a bit surprising that while the differences in 1992 were not very large, the average earnings among top earners were about 9 times higher than among the bottom 90%. This might be caused by after-transition factors. While before 1992 there were only government-owned firms in the country, now there are a number of international companies here. These companies can afford to pay higher wages, as they originate from the West and they are working in an environment where price levels are higher. So for them it is worth paying higher earnings compared to the inland companies, and this is still far lower than the earnings they pay in the countries they originate from. This might be a reason for the dynamic increase in earnings in the upper centiles of the wage distribution, while those in the lower part are more likely to work at some Hungarian firm which cannot afford such high wages. Also, these are the individuals who are less educated, so even if they work for an international company, they are in the kind of jobs where the difference in wages is not that wide.

Figure 3.2 shows the trends in average annual earnings in Canada, Sweden and the United Kingdom – the countries which are examined by Fortin et al (2017) and give the basis of the comparison of my results for Hungary.

Canada and Sweden have similar trends to each other. In both countries the trends in the different centile groupings seem to have the kind of dynamics as in Hungary. In the two lower centile groupings the average annual earnings are increasing only moderately, while there are huge increases in the upper part of the wage distribution. Surprisingly, however, the average annual earnings in the top 0.1% of the earnings distribution is lower in these countries than in the next 0.9% centile. This is quite controversial, as centile groupings were defined based on the wage distribution, meaning that the highest centile grouping should have the highest earnings.

3.3 Share of Women and Earnings Ratios by Centile Groupings

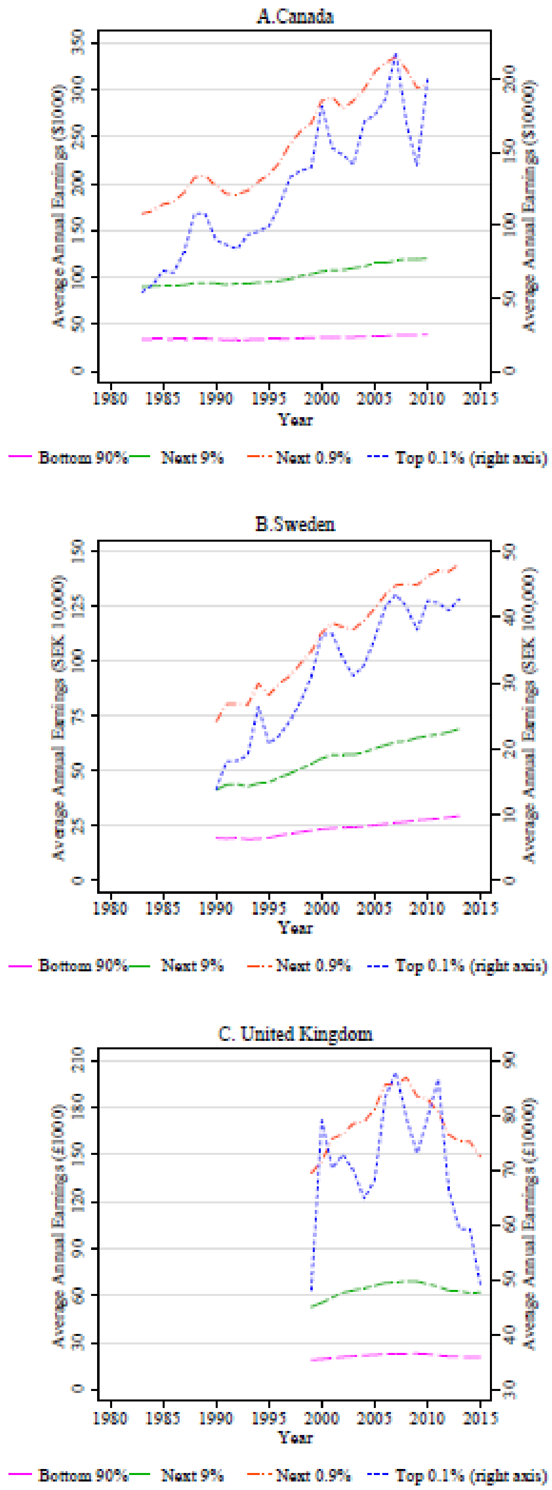
In this section I analyze the share of women in the four centile groupings of the wage distribution and the female / male annual earnings ratios by centile groupings, as well as the overall share of women and earnings ratios.

Figures 3.3 and 3.4 show the share of women and the female / male earnings ratios respectively.

It can be seen on figure 3.3 that the overall share of women in the labor force had a sudden increase in 1994 from 43% to 71%. In the following years, this share gradually decreased, dropping back to 65% in 2015. As expected, the share of women in the bottom 90% of the wage distribution is above the overall female share in labor force during these years starting from around the same level in 1992 and then going above 70% after the sudden increase in 1994. From then on, it was 2 percentage points above the overall share all along the years, on average.

The share of women in the next 9% of the wage distribution is much lower than the previous ones. It does not have the sudden change in 1994; it rather stays at around the same level at 50%, ending up at 51% in 2015 after some gradual increase. Unlike the female share in this centile grouping, the share of women in the next 0.9% has been decreasing over the years. It dropped to somewhat above 30% in 1994 and it gradually

Figure 3.2: Average annual earnings trends by centile groupings - Canada, Sweden, UK

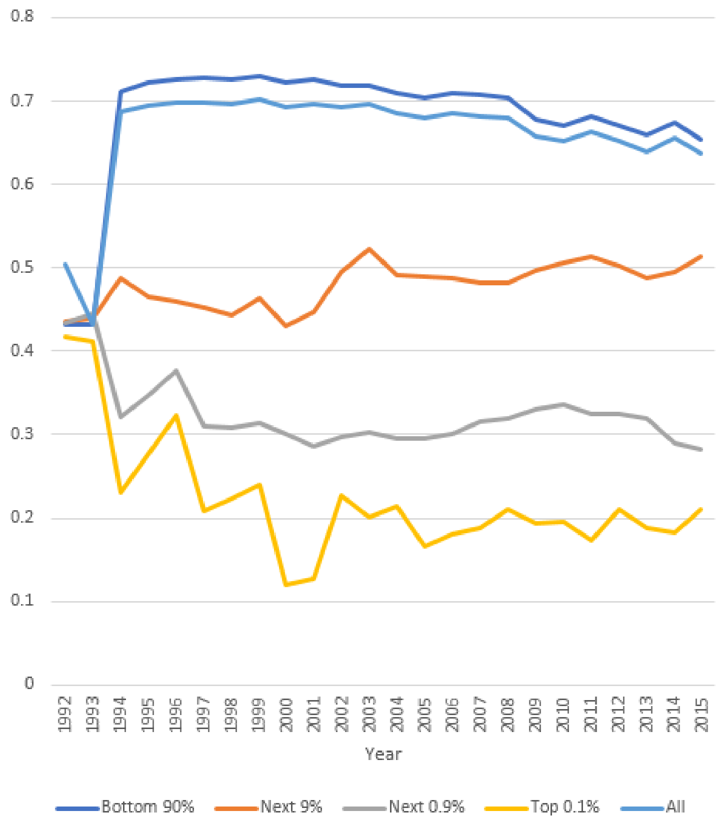


Source: Fortin et al. (2017)

decreased to 28% in 2015 after some changes around 30% beginning in the early 2000s.

According to Frey (1997), in the years after 1990 the employment rate had a sudden

Figure 3.3: Share of women by centile groupings

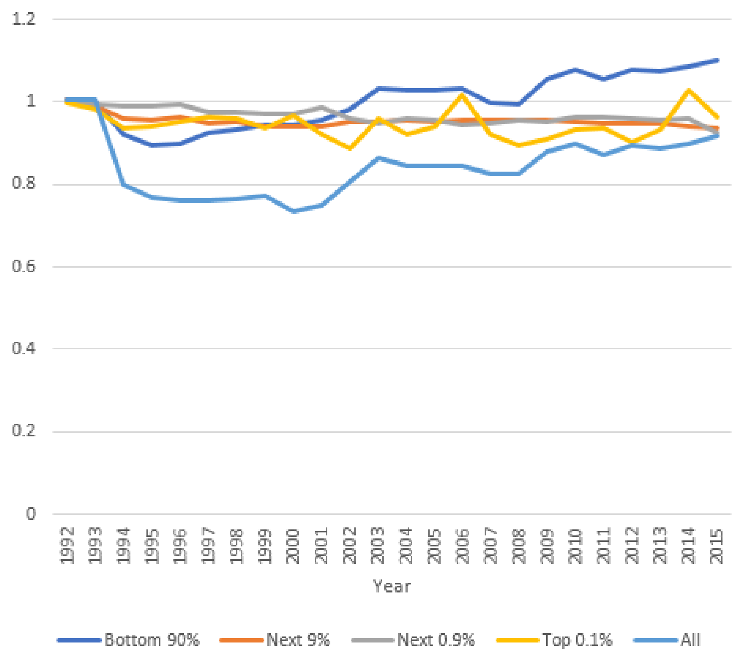


Source: Own calculations from the Hungarian wage survey data

drop due to the closing of several firms that employed lots of people in the ambition of reaching full employment. An even more serious problem was that two thirds of the people who became unemployed also became inactive at the same time. These changes were even larger among women. The share of women in the labor force reached that of the average level of the other members of the European Union, which does not sound bad in itself. However, unlike in the other countries, in Hungary, a higher share of women worked in part-time positions. The problem was still present, because the tendencies in Hungary were the opposite of those in the other countries in the EU.

The above mentioned changes might be the explanation for the sudden drop of the share of women in the labor force in the higher centile groupings of the earnings distribution. A more detailed potential explanation for this would be the following. When the sudden closing of lots of firms happened, those women in the top positions were in the highest danger of being thrown out of labor force. Those working in jobs that are needed

Figure 3.4: Female/male average annual earnings ratios by centile groupings



Source: Own calculations from the Hungarian wage survey data

every day to provide goods and services were in a better position, as their work was more essential. Even if they lost their job, it should not have taken much time to find a new one. However, those who were in the top positions, would have fewer opportunities until the entering of new, probably international firms in the Hungarian market, or the birth of new Hungarian firms.

On the other hand, Frey (1997) also mentions that it was mostly among those with lower education where extreme unemployment occurred. This is not quite in accordance with the above theory, but as the share of women in the labor force in the two middle centile groupings has not changed, it could happen that mostly those women lost their jobs who were not very highly educated, but had a position above what their education could explain. As Frey (1997) does not divide the wage distribution into centile groupings, there is no explanation that would tell what causes these trends for sure, while we have also no information about the connection between education and the position one had before the transition.

Although the change in the top 0.1% of the distribution still seems to be relatively small, it is important to highlight that the earnings trend does not match the trend in the share of women in the labor force and among top earners. While average earnings are rapidly increasing in this part of the distribution, the share of women has a decreasing trend over these years.

Figure 3.4 shows the average annual female/male earnings ratios by centile groupings, as well as the overall gender pay ratio. As can be seen on the graph, there are not as many differences in the earnings ratios of the different centile groupings as there are in the share of women. In all groups, the earnings ratio started at around 1 in 1992, meaning that the average earnings in all centile groupings, an overall in the distribution, was the same among men and women. Then, in 1994 the overall ratio dropped to 80%, so men started to have higher average earnings when considering the whole distribution. After a gradual increase, it reached 91% in 2015. As this is the lowest line in the graph, we can already conclude that the average earnings ratio is less favorable when taking into account the whole distribution.

The earnings ratio in the bottom 90% of the distribution has a very similar trend to that of the overall share, having a smaller drop in 1994 and then going above 1 in 2003. In 2015 it was at the 1.1 level, being the highest earnings ratio among the centile groupings.

The rest of the centile groupings do not show large changes in the earnings ratio. They all fluctuate around the 95% level, meaning that men earn somewhat higher earnings than women on average. The top 0.1% centile is different from the other two only in volatility. Although fluctuating around the 95% level like the others (except the bottom 90%), the average earnings ratio in this group was less than the others in several years, and it even went over the 100% level in two years. In this group the ratio ended up at the 96% level in 2015, while the other two were at 93%.

According to this figure, it seems that the trend change in the average annual earnings ratio taking the whole distribution was mainly driven by the change in the earnings ratio in the bottom 90% of the wage distribution, while the earnings ratio in the other parts of the distribution remained more or less the same in the last 20 years. Given that the hours worked are not available in the data set, the reason for these trends is unclear. There are two main explanations for the phenomenon. First, it can happen, that women in the bottom 90% of the distribution might start to work more hours, increasing their annual earnings this way. Another explanation can be that equality increased in this part of the wage distribution, and earnings per hour got closer among men and women. A third, yet unlikely, explanation would be that male earnings got lower earnings while female earnings rose or did not change.

The drop in the average earnings ratios in the 1990s can be taken as natural and expected after the transition. Later, the growth of the earnings ratio can also be taken as expected and might be attributed to international companies entering the country and offering more equality to women in terms of positions and earnings as well. However, the graph does not seem to support this theory, as the earnings ratio remained the same in the top three centile groupings of the earnings distribution.

Another explanation might be found in the minimum wage laws that were introduced during these years. These laws can explain the increase of the earnings ratio in the lowest centile grouping, while its relative invariance in the other centile groupings, thus having the lowest centile grouping to drive all the change that happened in the overall earnings ratio.

According to KSH, the minimum wage in 1992 was 35.8% of the average gross earnings. Since then this number has grown almost every year. In 2015 it reached 42.2% of the gross average earnings. This seem to support the above theory saying that the increasing earnings ratio in the bottom 90% of the distribution might be explained by the changes in the minimum wage. Although this is only about the wage levels, if we take into account that women might have jobs that pay less, it follows that the changes in the minimum wage change their situation more than that of men, tending to rise in the earnings ratio.

These figures do not support the theory about a strong glass ceiling effect in Hungary. Even if the share of women in the top 0.1% of the wage distribution is much lower than in the other centile groupings, the earnings ratio is not significantly different from that of the other centile groupings, though more volatile.

Figure 3.5 shows the female shares and gender ratio in average annual earnings by centile groupings for Canada, Sweden and the UK.

The share of female had very similar trends in the three countries shown on the figure. In all three countries, the share of women is lowest in the highest centile grouping and it

Figure 3.5: Female shares and gender ratio in average annual earnings by centile groupings - Canada, Sweden, UK

Source: Fortin et al. (2017)

gets higher as taking lower and lower centile groupings. Also, the share of women is higher than the overall share only in the bottom 90% of the wage distribution. Furthermore, the share of women has an increasing trend in the three upper centile groupings, while it is more or less constant or very slightly increasing in the bottom 90% of the earnings distribution and overall.

Compared to figure 3.3 we can see that, even the order of the lines belonging to each centile grouping and the overall distribution is not the same as in Hungary, where the share of women did not increase after the sudden change in 1994 in any of the centiles. Contrariwise, it is decreasing in the two top centile groupings of the distribution, the bottom 90% and overall, while it stays around a constant level in the middle two centile groupings. This difference in trends suggest some basic and maybe hereditary differences between the three western countries and Hungary. This may be for a number of reasons like being in different stage of development in the 1990s, the three western countries having some benefits already when Hungary has only stepped on the path of capitalist development.

Figure 3.5 also shows the gender ratios in each centile grouping. While there are no major changes in the three centile groupings belonging to the lower part of the wage distribution, the gender ratio in the top 0.1% of the wage distribution decreased over time in all three countries. These trends are very different from those seen in Hungary, as the gender ratios were around some constant value in all centiles but the bottom 90%, where the ratio has increased over time, driving the increasing trend of the gender ratios in the whole distribution.

In the next chapter I start by counting counterfactuals to get some more insight about how would the earnings ratio look like if the share of women in each centile groupings would be the same as the share of men, that is, if there was no gender share inequality to influence the earnings ratios.

Chapter 4 – Empirical Analysis

In this chapter I present the main empirical analysis of my thesis. I start with the technical details of the counterfactual analysis. Then, I present the *OLS* regressions which are the methodological tools for doing this analysis. Finally, I use the Oaxaca-Blinder decomposition method to learn how much of the gap is due to the inequality, and I detail and interpret the results.

4.1 Counterfactual Analysis of Earnings Ratios

Following Fortin et al (2017), the main questions of the counterfactual analysis are: "What would be the average wage of women if they had the same characteristics as men?" and "What would be the average wage of men if they were paid as women?"

To answer these questions we can use *OLS* regressions of the following form

$$Y_i^g = X_i \beta_g + u_i^g$$

where Y_i^g is earnings, X_i are individual characteristics, and $g = 0, 1$ stands for gender groups. This way, we get gender-specific regressions of the above form. Given the *OLS* method, we have $\mathrm{E}[u_i^g \mid X_i] = 0$, as the exogeneity of the explanatory variables is required. This means that we assume that the regressors are not correlated with the error terms. This is a very strict, but necessary assumption, as otherwise the *OLS* estimator would be biased.

Let us define a dummy variable for gender group of the following form

$$D_i = \begin{cases} 1, & \text{if man} \\ 0, & \text{if woman} \end{cases} .$$

Then the overall mean wage gap can be written in the following form

$$\Delta = \mathrm{E}[X_i \mid D_i = 1]\beta_1 - \mathrm{E}[X_i \mid D_i = 0]\beta_0,$$

where the first term is called the unexplained part, while the latter one is called the explained part which is due to differences in characteristics X_i.

According to Kline (2011), this counterfactual can either be computed with regression coefficients or by re-weighting proposed by DiNardo et al (1995). This way, we take the share of women in each centile grouping of the earnings distribution and make it equal to the share of men in that given centile. Then, we re-weight the wages with the weights we created in order to get simulated wage values for a case when the share of women equals that of men. This can also be imagined by distributing women equally through the whole wage distribution, that is, rearranging women from the lower centile groupings into the upper part of the distribution, as the share of women is far higher in the lower tail of the earnings distribution than in the top centiles.

For re-weighting, we need the propensity score $P(X_i) = P(X_i \mid D_i)$ and the probability of an individual belonging to the gender group 1 (males) $\pi = P(D_i = 1)$. Using these expressions, the re-weighting factor can be written in the following form

$$w_i(X) = \frac{P(X_i)}{1 - P(X_i)} \frac{1 - \pi}{\pi}$$

In the sample, this can be calculate simply by gender ratios in the different groups we define

$$\hat{w}_i = \frac{N_{1j}}{N_{0j}} \frac{N_1}{N_0}, \tag{4.1}$$

where N_{gj} denotes the number of individuals belonging to groups $g = 0, 1$ in group j, while N_0 and N_1 denote the overall number of males and females in the sample.

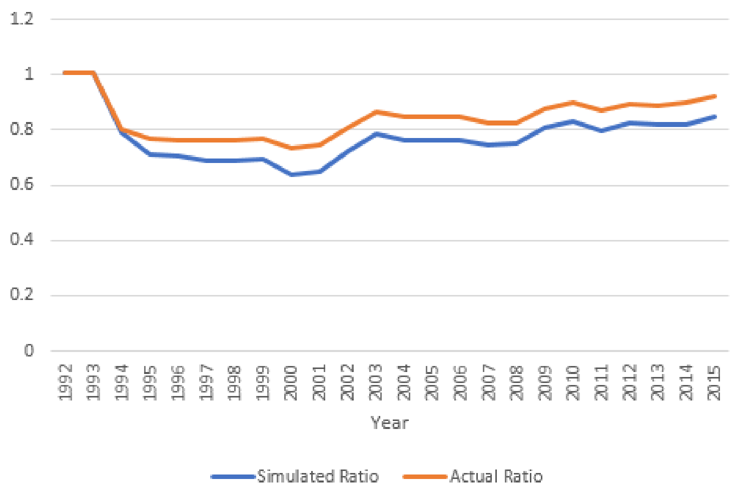
As I would like to build the counterfactual in a way that gives me the earnings ratios cleaned from the effect of gender inequalities in the different centile groupings of the earnings distribution, I use these centile groupings as the groups j when calculating the re-weighting factor as defined in equation 4.1.

This way, the counterfactual tells us "what would be the female/male earnings ratio if the share of women in the centile groupings of the wage distribution was the same as that of men?". We could also use other category variables like education, age, etc. Then, the counterfactual would allow us to know what the female/male earnings ratio would be if the human capital of women was on the same level as that of men in each centile grouping of the earnings distribution.

I calculated the female/male average annual earnings ratios in each year using this re-weighting factor to control for the unequal gender shares in each centile grouping, disregarding the human capital considerations this time.

Figure 4.1 shows the results of the counterfactual calculations, as well as the actual female/male average annual earnings ratios.

Figure 4.1: Simulated female/male average annual earnings ratios



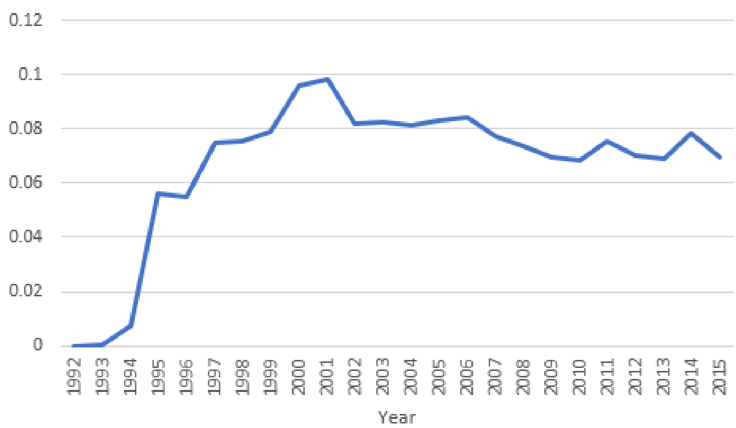
Source: Own calculations from the Hungarian wage survey data

Following the discussion of figure 3.3, results shown in figure 4.1 should not be surprising. The share of women has not increased in most centile groupings, especially not on the top of the earnings distribution, thus the simulated earnings ratio should not be more favorable than the actual one. Indeed, except for 1992-1994, where it moves together with the actual earnings ratio, it is constantly below the actual ratio in the following years. On average, it is about 7 percentage points lower than the actual ratio, ending up at the 85% level in 2015, while the actual ratio was 92% in this year. This means that if the share of women in our centile groupings would be the same as the share of men, then earnings inequality would rise instead of shrinking. This implies that even if the share of women in the top part of the distribution is less than the share of men, and it is far below the share of women in the other parts of the earnings distribution, having the same share as that of men, women would be worse off, rather than better off.

Furthermore, figure 4.1 shows that the actual earnings ratio has improved over time. After a pretty significant drop in the 1990s, it started to get closer to the 100% level.

Figure 4.2 shows the difference between the actual and simulated gender earnings ratios.

Figure 4.2: Difference between actual and simulated female/male average annual earnings ratios



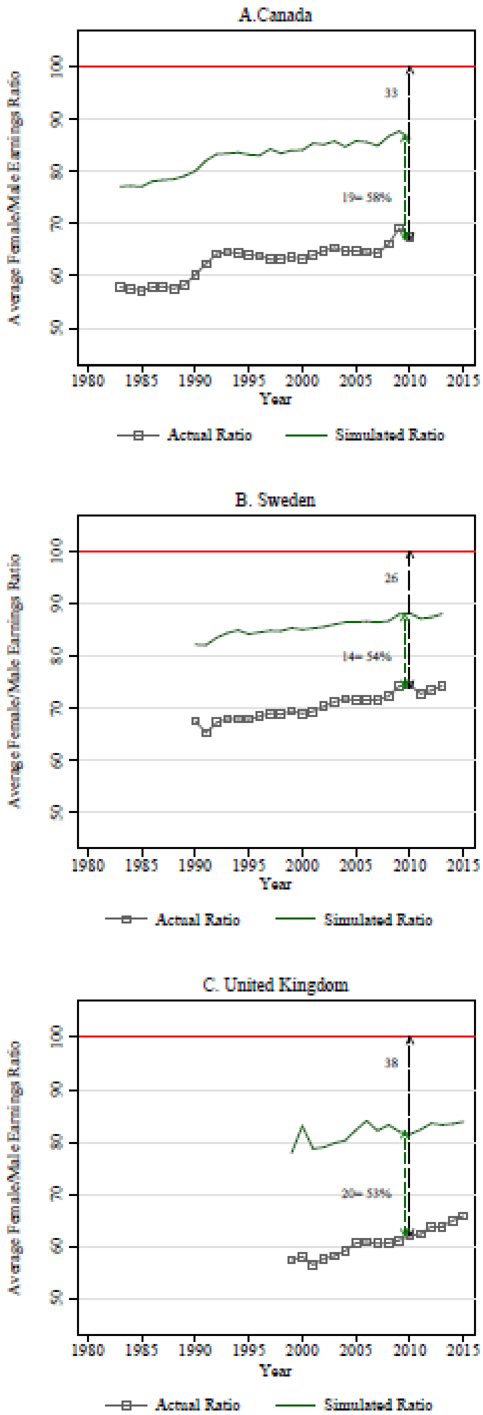
Source: Own calculations from the Hungarian wage survey data

Compared to the 1992 level when the two ratios are roughly the same, there has been some deterioration in the earnings ratio, meanings that the improvements in the under-representation of women in the top 0.1% of the earnings distribution has not kept pace with the increases in average top earnings.

Figure 4.3 shows the counterfactual gender ratios in average annual earnings in Canada, Sweden and the UK.

As the share of women has an increasing trend in all three countries, it is not surprising that setting the share of women to be equal to the share of men in each centile grouping makes the simulated average annual earnings ratios higher than the actual one. In these countries, the difference between the actual and the simulated ratio is somewhere between 15 and 20 percentage points, meanings that gender equality in the centile groupings would be very favorable for women in terms of earnings ratios.

Figure 4.3: Counterfactual gender ratios in average annual earnings ratios - Canada, Sweden, UK



Source: Fortin et al. (2017)

On the other hand, Hungary shows differences from these countries right away when checking the basic trends in the variables of interest. Unlike in the western countries in my comparison, the share of women in the top 0.1% of the earnings distribution has decreased significantly over time and it has not changed much in the other centiles. Thus, it is not surprising what the simulated earnings ratios show. Instead of having higher simulated earnings ratios, they are lower in the Hungarian case, meaning that while in the other countries equality would be favorable, it seems that this is not the case in Hungary.

4.2 Oaxaca-Blinder Decomposition

I use the Oaxaca-Blinder decomposition to learn how much of the overall male/female gender wage gap the different factors account for. I do the decomposition for the years 2006 and 2015. The former was chosen as one of the two years which are still before the great recession, but after major changes in the minimum wage in 2000 and 2003. The latter one was chosen based on considerations of relevance. I use two nested models. Model 1 includes the centile groupings, demographics (age only), education and region only, while Model 2 adds industry and occupation.

The above mentioned variables are defined in the following way. Age is in the range 25-64 and it is divided into 8 five-year bins. In the data set education is given by qualification attained and years of education is defined based on the years in school each qualification requires. Given this feature of the data, I use the qualification dummies as a measure of education. These dummies define the following categories: unskilled workers, skilled workers, high school graduates and university or collage graduates. The region variable takes on 20 values in both years corresponding to the counties of Hungary and refers to the place of work. I use 17 industry categories in 2006, while the data set contains 21 categories of this variable in 2015; these are also defined by dummy variables. Occupations are given based on the FEOR categories (Uniform Classification System of Occupations). I use the 42 main categories of occupations defined in the system, thus I code all four-number identifiers based on its first two numbers, which refer to the largest and second largest occupation classes.

Table 4.1 shows the results of the Oaxaca-Blinder decomposition for both models and years.

When using Model 1, 66.53% of the gap is explained overall in 2006, which is decom-

Table 4.1: O-B decomposition of gender differentials in annual earnings

Explanatory variables	Model 1				Model 2			
	2006	% of gap	2015	% of gap	2006	% of gap	2015	% of gap
Raw Gender Earnings Gap	26419.36		18027.99		26419.36		18027.99	
Accounted for by differences in characteristics								
Centile Groupings	23135.57	87.57	23478.05	130.23	21665.10	12.49	22142.70	122.82
Demographics (age only)	-1118.54	-4.23	-1794.87	9.96	-1152.23	-4.36	-1469.62	-8.15
Education	5115.43	-19.36	16514.32	-91.60	4090.05	-15.48	7092.46	-39.34
Region	674.61	2.55	-35.55	-0.19	598.62	2.27	69.93	0.39
Industry					963.16	3.65	7333.04	40.68
Occupation					4495.62	17.02	-13968.91	-77.48
Total Explained	17576.19	66.53	5133.29	28.47	19209.54	72.71	7014.68	38.90
Total Unexplained	8843.17	33.47	12894.69	71.52	7209.82	27.28	11013.31	61.09

posed the following way. The four centile groupings account for 87.57%, while age and education explain negative amounts: -4.23% and -19.36% respectively. Region accounts for 2.55% of the gap. These results show, first of all, that women are more educated than men are, thus discrimination is higher than the difference in wages.

I got similar results for 2015. However, here, not only that the share of the gap explained by the centile groupings is nore than the total explained share, but also it is more than the overall gap: 130% of the gap is explained by the centiles. Age and education account for negative amounts of the gap in 2015 as well, although education accounts for -90%, almost a whole negative gap. Here, region has a negative effect too.

More specifically, as one form of discrimination can be in hiring/ not hiring a given individual, the positive effect of the centile groupings show that there are more men in the top positions than women, thus giving a very positive gap, ceteris paribus. In terms of education, a negative share of the male/female gap being explained by this variable, women are more educated than men which makes the gap we examine negative, everything else held constant.

A similar argument holds in case of the age groups. The defined age bins have a negative explanatory effect on the gap, meaning that women are usually in a higher share than men in most of the age groups.

Similar to Fortin et al (2017), centile groupings in Hungary also explain a positive share of the male/female wage gap. However, while Model 1 explains more than 87.57% of the gap in 2006 and 130% in 2015 in Hungary, this share is only 17.9% in Sweden in 1997, while it is 16.5% in 2015. Also, while adding the industry and occupation variables to the model decreases the share of the gap explained by the centile groupings in Hungary in both years, to 12.5% and 122% respectively, in case of Sweden, there is a consistent growth in both years, the share being 42.4% and 37.1% respectively.

The Mincer-Polachek hypothesis can be rejected in the case of Hungary too. Increases in education for women does not imply increases in earnings, shown by the effect of education being negative.

Like in Fortin et al (2017), the addition of covariates has also reduced the explanatory power of the centile groupings. However, while in the case of Sweden the explanatory power of the centiles remains overwhelming, in case of Hungary it diminishes to 12% in 2006, while it remains overwhelming in 2015.

Another similarity to the results of Fortin et al (2017), the explanatory power of the centile groupings has also increased over time in case of Hungary. Thus, we can drive the same conclusion: top earnings inequality is becoming an important "‘swimming upstream"’ factor.

Although Fortin et al (2017) also did the same analysis for Canada and the UK, I avoid comparisons with these countries, because my data is less similar to theirs than that of Sweden. For example, they can control for more demographic variables, and they have hourly wage as a better measure of earnings. Another difference is that given survey data, they do not have as many people from the top tail of the earnings distribution as for Sweden. Hence, comparison would perhaps point out these basic differences rather than real differences between the results in these countries.

Chapter 5 – Conclusion

This thesis has studied the connection between the gender inequality among top earners and the overall gender pay gap. I did this by dividing the earnings distribution into four, mutually exclusive, centile groupings in the following way: bottom 90%, next 9%, next 0.9% and top 0.1%. I borrowed this division from the literature, as well as following the decision of Fortin et al (2017). Using these centile groupings I first analyzed the female shares and actual annual earnings in each part of the distribution. Then, I built counterfactuals that control for the differences in gender shares in each of the four centile groupings, resulting in a simulated wage distribution assuming the same share of women that the actual share of men is in each of the centile groupings.

In my analysis I found that the overall share of women in the labor force, as well as the share of women in the top 0.1% of the earnings distribution, have decreased over the past 23 years, while the average annual net earnings has increased.

At the same time, the average annual female/male earnings ratios have increased in the sample overall, as well as in the bottom 90% of the wage distribution. The ratio has fluctuated at the same level in the other centile groupings.

The counterfactual analysis gives some useful insights about the average annual female/male earnings ratios, given that the share of women in each centile groupings was the same as the share of men. An interesting finding here is that unlike what one would have expected, the simulated earnings ratio is lower than the actual one all over the years studied, meaning that women were worse off if their share in the given centiles of the distribution was the same as that of men.

To get more precise estimates on the explanatory power the centile groupings and other factors on the gender wage gap, I used the usual Oaxaca-Blinder decomposition as a tool. I did the calculations for two years and two different models. This way, I found that while in case of the model including the centile groupings, age, education and region only, the the centile groupings account for a huge share of the gap being even above 100% in 2015, in case of the wider model including industry and occupation, the part of the gap ecplained by the centiles drops way below its value in the other model in 2006, while it shows some decrease also in 2015 however, it still remains above 100%).

Compared to the results of Fortin et al (2017) for Canada, Sweden and the United Kingdom, the case of Hungary seems to be very different right away from the beginning of the analysis. First, the share of women in labor force, but especially in the top 0.1% of the earnings distribution changed the opposite way to the three countries of comparison. Then, after some differences discussed in the descriptive statistics part, there is a very significant difference in the empirical results. While for the three countries analyzed by Fortin et al (2017) the simulated earnings ratios give the results one could have expected showing that if the share of women in each centile grouping was equal to the share of men, then there would be huge increases in the earnings ratios. On the contrary, the same analysis on the Hungarian data gives that if the share of women was the same as that of men, women would be worse off and the earnings ratio would be lower than the actual one.

Finally, unlike the for the three countries of the other study, the inequality turns out to be larger than the difference in earnings in Hungary, shown by the centile groupings accounting for more than 100% of the gender wage gap in the Oaxaca-Blinder decomposition.

As a weakness of the analysis, it has to be mentioned that the hours worked was not available in the data set, thus making the analysis a bit less accurate, as we could not distinguish between wage differences caused by different hours worked and those caused by unequal treatment. Furthermore, individual weights could be used in later analyses in order to get more precise estimates of the averages calculated, however, this should not make much of a difference given that the sample is large in each year. Moreover, including more demographic variables in the decompositions would have raised the quality and accuracy of the results.

Bibliography

Albrecht J., Björklund A., Vroman S. (2003): Is there a glass ceiling in sweden? *Journal of Labor economics* **21**(1), 154–177.

Alvaredo F., Atkinson A., Pikkety T., Saez E (2013): The top 1 percent in international and historical perspective. *Journal of Economic perspectives* **27**(3), 3–20.

Arulampalam W., Booth A., Bryan M. (2007): Is there a glass ceiling over Europe? exploring the gender pay gap across the wage distribution. *ILR Review* 60(2), 163–186.

Atkinson A. B., Casarico A., Vitchocsky S. (2016): *Top incomes and the gender divide.* Melbourne Institute Working Paper No 27/16.

Baker M., Drolet M. (2010): A new view of the male/female pay gap. *Canadian Public Policy* **36**(4), 429–464.

Bayer P., Charles K. K. (2016): *Divergent paths: Structural change, economic rank, and the evolution of black-white earnings differences 1940-2014.* (No w22797) National Bureau of Economic Research

Blau F. D., Kahn L. M. (1997): Swimming upstream: Trends in the gender wage differential in the 1980s. *Journal of labor Economics* **51**(1), 1–42.

Blau F. D., Kahn L. M. (2017): The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature* **55**(3), 789–865.

Brewer M., Sibieta L., Wren-Lewis L. (2008): *Racing away? income inequality and the evolution of high incomes.* Institute for Fiscal Studies Briefing Note 76

Cukrowska-Torzewska E., Lovász A. (2016): Are children driving the gender wage gap? *Economics of Transition* **24**(2), 259–297.

Davies M. L. (2015): Improving the gender balance on British boards: Women on boards 5 year summary. *BIS research papers and Corporate governance, Department of Business, Innovation & Skills*, London

DiNardo J., Fortin N. M., Lemieux T. (1995): *Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach.* (No w5093) National bureau of economic research

Edlund L., Kopczuk W. (2009): Women, wealth, and mobility. *American Economic Review* **99**(1), 146–178.

Fortin N. M., Lumieux T. (1998): Rank regressions, wage distributions, and the gender gap. *Journal of Human Resources*, 610–643.

Fortin N. M., Lumieux T. (2000): Are women's wage gains men's losses? a distributional test. *American Economic Review* **90**(2), 456–460.

Fortin N. M., Bell B., B́ohm M. (2017): Top earnings inequality and the gender pay gap: Canada, sweden and the united kingdom. *Labour Economics* **47**, 107–123.

Frey M. (1997): Nök a munkaeröpiacon. In: *Lévai K, Tóth IG (eds) Szerepváltozások. Nök és férfiak a munkaeröpiacon, 1997*, TÁRKI, Munkaügyi Minisztérium Egyenlö Esélyek Titkársága, 13–34.

Gabaix X., Landier A. (2008): Why has ceo pay increased so much? *The Quarterly Journal of Economics* **123**(1), 49–100.

Gneezy U., Pietrasz A. (2013): *When half of the men are more competitive than all women.* Unpublished manuscript

Guvenen F., Kaplan G., Song J. (2014): *The glass ceiling and the paper floor: Gender differences among top earners, 1981-2012.* (No w20560) National Bureau of Economic Research

Kline P. (2011): Oaxaca-blinder as a reweighting estimator. *American Economic Review* **101**(3), 532–537.

Lovász A. (2014): *Do women have better opportunities in the public sector? An analysis of the gender pay gap and occupational segregation in the public and private sector, The Hungarian Labor Market 2014*

Lovász A., Telegdy Á. (2010): Labour market discrimination - types, measurement issues, empirical solutions, the hungarian labour market - review and analysis 2010. *IE-HAS, OFA*, 46–67.

Lovász A. (2016): Childcare expansion and mothers' employment in post-socialist countries. *IZA World of Labor*

Mincer J., Polachek S. (1974): Family investments in human capital: Earnings of women. *Journal of political Economy* **82**(2), S76–S108.

Pikkety T., Saez E. (2003): Income inequality in the united states, 1913-1998. *The Quarterly journal of economics* **118**(1), 1–41.

Pikkety T., Saez E. (2013): Top incomes and the great recession: Recent evolutions and policy implications. *IMF economic review* **61**(3), 456–478.

Pikkety T., Saez E., Statntcheva S. (2014): Optimal taxation of top labor incomes: A tale of three elasticities. *American economic journal: economic policy* **6**(1), 230–271.