Gender equality: remarkable improvements in education but still a long way to go in the job market*

Wenxuan Li

21 February 2022

Abstract

Gender equality is a significant topic for everyone, not only female, but also male. In this paper, we summarize the current status of the gender equality in the education field and in the job market. We find that there is a remarkable improvement in the education – female have over-taken education than male, however, things are totally different in the job market. We include a brief discussion on the reasons for this phenomena referencing the existing literature, and hope it can help the gender equality in the job market in the future.

1 Introduction

Gender equality is always an important topic across the whole world. It influences not only female, but also male. There are a bunch of literature talks the cause, the results, and the impact of gender inequality, but there is no paper summarizing the current gender equality status. So what is the current status? Is there any improvement is the past several years? What should we do in the next few years? In this paper, we will try to answer these questions.

We adopt a combination of several data sets to find the current status for gender equality from the education and the job market point of view. We find there is a great improvement in the education – more and more female have the opportunity to attain a Bachelor degree or higher and choose their majors in STEM. However, when things come to the job market, it is a totally different story – male still participate more than female and earning higher salaries. This may be caused by the motherhood penalty and the stereotype put on female. If we want to relieve the gender inequality in the job market, the policy maker may want to cut in from these two aspects.

The remaining content is organized as follows. Section 2 introduces the data set we adopt in this paper. Section 3 presents the results based on the data set. Section 4 connects the results to the gender equality and link them to the literature to make a broader dicussion.

2 Data

In this paper, the final data set we adopt has come through multiple cleaning processes. Instead of the original survey data set, we adopt the pre-processed data set from Bertrand (2020). By pre-processing, we mean collect, combine, and organize different data sets to a panel data set for each source. The paper, Bertrand (2020), provides us with three panel data sets: one data set for each data source, including OECD OECD (2022), world bank Bank (2022), and ILO ILO (2022). These panel data sets are still too large to find useful information, which requires further cleaning processes – the OECD data set contains 2622 samples with 1875 variables; the world bank data set contains 12803 samples with 1895 variables; and the ILO data set contains 10382 samples with 6188 variables. Due to the size of data sets, we temporarily do not introduce them in

*Code and data are available at: https://github.com/leoli2022/sta304/tree/main/paper2. This is a replication work of paper DOI: 10.1257/pandp.20201126.

details. Further cleaning processes are conducted, including filtering appropriate samples, selecting useful variables, combining different data sets by two variables, country and year, and formulating a final data set. We refer this data set as processed data set in the following content.

2.1 Variable and explanations

Some basic information of the processed data set is discussed as follows. It contains 324 observations on 23 variables. Since we are combining three data sets to the final one, the processed data set contains some non-answer cells. The processed data set consists of variables as shown in Table 1. The first two variables serve as the "ID" of the sample. The third variable to the 10-th variable are for the share of different genders who attain at least a bachelor degree among different age groups. The 11-th to the 20-th variable are for the share of different gender who major in STEM majors among different education level. The 21-st variable is for the log value GDP per capita. The 22-nd to 23-rd variable are the labour force participation of different genders. These variables will support us to conduct a basic analysis on the gender inquality in the recent year.

Table 1: Variables and their meaning

Variable Name	Explanation
`ctry_iso`	ISO code of country
`year`	Year
`female_bach_2534`	Shares of female with at least a bachelor's Degree between 25 and 34
`female_bach_3544`	Shares of female with at least a bachelor's Degree between 35 and 44
`female_bach_4554`	Shares of female with at least a bachelor's Degree between 45 and 54
`female_bach_5564`	Shares of female with at least a bachelor's Degree between 55 and 64
`male_bach_2534`	Shares of male with at least a bachelor's Degree between 25 and 34
`male_bach_3544`	Shares of male with at least a bachelor's Degree between 35 and 44
`male_bach_4554`	Shares of male with at least a bachelor's Degree between 45 and 54
`male_bach_5564`	Shares of male with at least a bachelor's Degree between 55 and 64
`female_l5t8`	Shares of female with at least a tertiary degree in STEM
`female_l5`	Shares of female with a Tertiary degree in STEM
`female_l6`	Shares of female with a Bachelor degree in STEM
`female_l7`	Shares of female with a Master degree in STEM
`female_l8`	Shares of female with a Doctor degree in STEM
`male_l5t8`	Shares of male with at least a tertiary degree in STEM
`male_l5`	Shares of male with a Tertiary degree in STEM
`male_l6`	Shares of male with a Bachelor degree in STEM
`male_l7`	Shares of male with a Master degree in STEM
`male_l8`	Shares of male with a Doctor degree in STEM
`log_gdp`	The log value of GDP per capita
`female_labour`	Female labour force participation
`male_labour`	Male labour force participation

2.2 Visualizations

In this subsection, we are going to plot some variables, which are closely related with gender. Actually, all variables except for country code, year, and GDP are visualized. Figure 1 shows two histograms for the share of people who attain at least a Bachelor degree among different age groups for two genders. From the plot, we can see that in age group 45~54 and 55~64, less share of female tends to attain a bachelor or above degree. But the share of male tends to be quite flat. The distribution of the share does not change much among different age groups.

Figure 2 shows the histogram for the share of people who major in STEM majors across different education level in different gender groups. From the plot, we can straightforwardly feel that a much larger share of

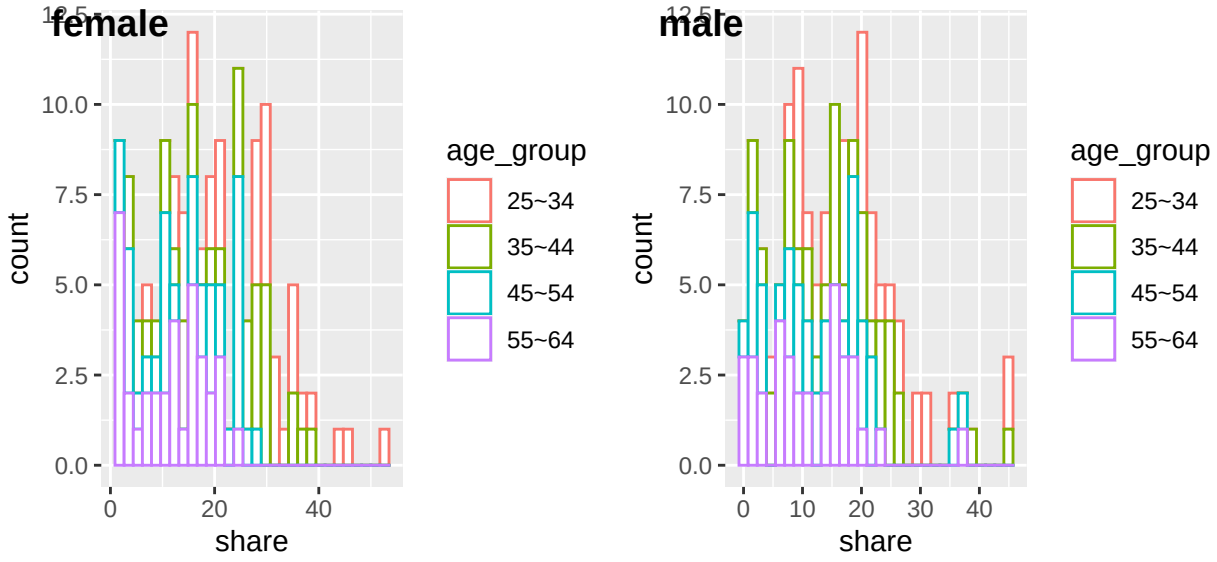

Figure 1: Histogram for the share of people who attain at least a Bachelor degree from different genders

male receive STEM education rather than other majors. But as the education level increasing (i.e., from level 5 to level 8), the different is becoming smaller and smaller.

Figure 3 is a histogram to quantify the labour participation of the female and male. From the figure, we can clear see that male labour participation samples from different countries are close to 100%, however, the female labour participation centralizes between 50% to 90%.

3 Results

In the beginning of the presentation for the results, we want to first emphasize the analysis tool we adopt in this paper. We run our analysis in `R` (R Core Team 2020) and Rstudio (RStudio Team 2020). Several libraries are utilized in process, including `tidyverse` (Wickham et al. 2019), `ggpubr` (Kassambara 2020), `haven` (Wickham, Miller, and Smith 2022), `here` (Müller 2020), `ggplot2` (Wickham 2016), `ggrepel` (Slowikowski 2021), and `vtable` (Huntington-Klein 2021).

3.1 Summary statistics

In this section, we show the summary statistics of all variables except country code and year. Table 2 shows the summary statistics for variable of interest. From the table we can see that even though the data set contains 324 samples, most variables only have ~30 valid samples, which has been explained in the previous context. If we focus on the mean value of these variables, we can see that there is a clear increasing trend for the share of female who at least have a bachelor degree with the age group going down. The male shows similar trend, but with a flatter rate. As for the STEM major, no matter the education level is, a higher proportion of male shows up than female. This largest difference centers on the Tertiary: the share of male is more than 4 times of the female.

3.2 Some interesting figures

In this section, we will use some figures to show the interesting aspects of the data set. Unlike what we have done in Section 2, we will use more format of figures. Figure 4 is a a group of four scatter plots showing the share of female with bachelor's degree or higher versus the share of male in four age groups. One straight line

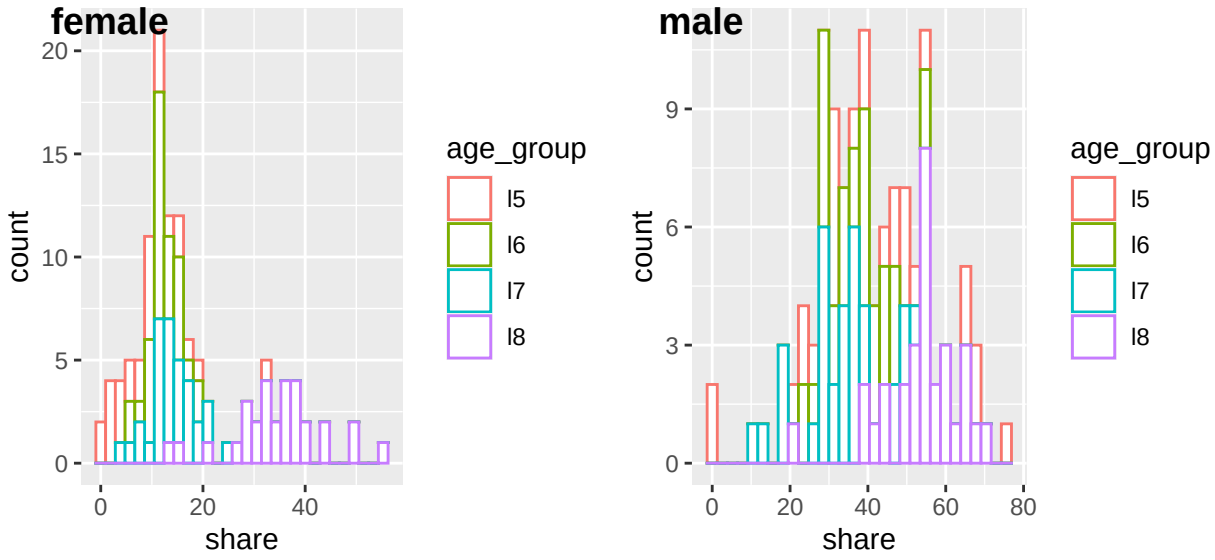

Figure 2: Histogram for the share of people who major in STEM from different genders

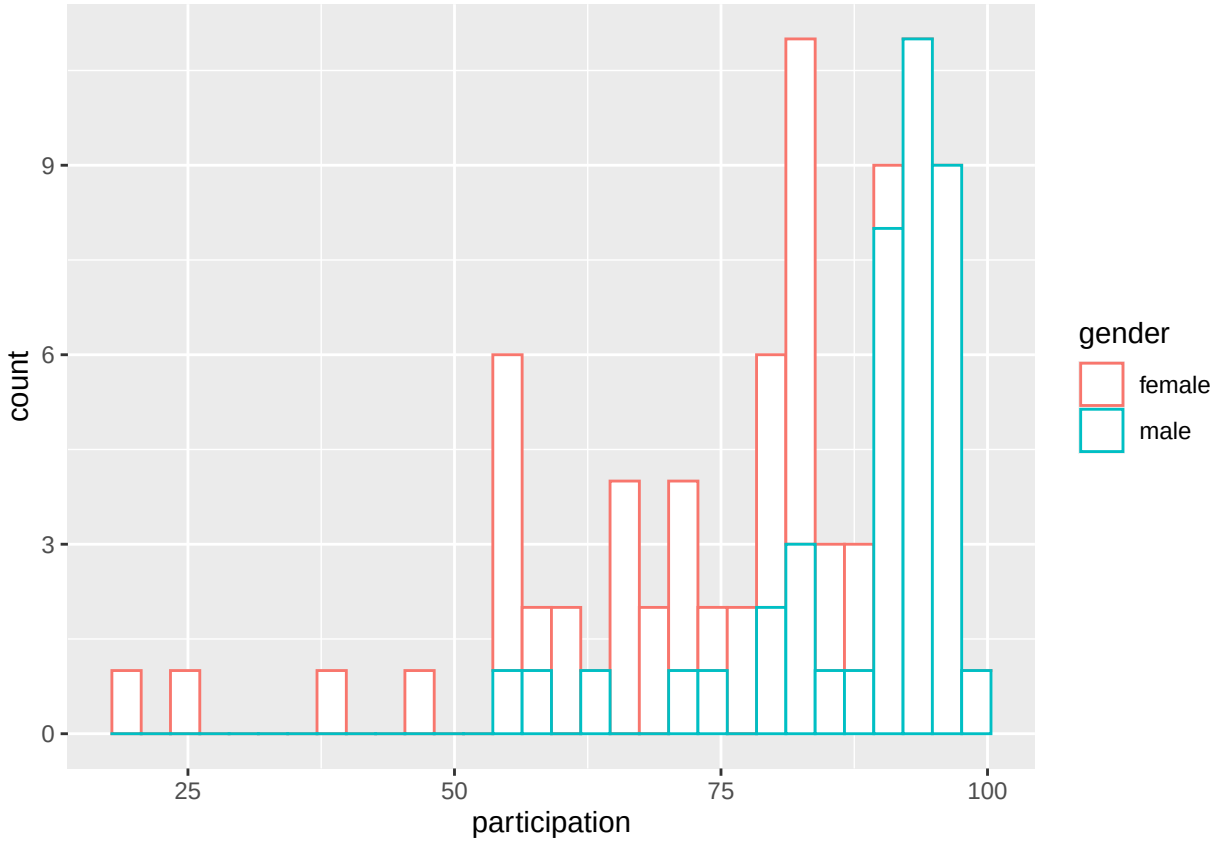

Figure 3: Histogram for labour participation

Table 2: Summary Statistics

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
female_bach_2534	35	26.547	10.515	7.211	17.696	32.893	51.797
female_bach_3544	35	21.114	10.464	2.859	13.132	28.888	39.104
female_bach_4554	35	15.272	8.281	1.567	9.74	22.902	27.615
female_bach_5564	35	11.401	6.982	0.951	4.536	16.709	24.053
male_bach_2534	35	20.192	9.824	4.756	12.798	25.695	45.203
male_bach_3544	35	17.066	9.77	1.289	10.106	23.29	45.447
male_bach_4554	35	13.098	9.126	0.622	6.964	18.839	36.651
male_bach_5564	35	11.262	8.033	0.519	5.699	16.619	36.437
female_l5t8	29	12.934	3.046	6.485	11.143	14.814	19.352
male_l5t8	29	38.699	7.243	25.52	35.205	43.322	53.394
female_l5	27	8.845	6.903	0	4.108	11.169	31.578
male_l5	27	40.721	19.177	0	30.181	50.962	75.591
female_l6	31	12.204	3.347	6.165	10.346	14.282	19.659
male_l6	31	37.382	8.368	22.743	30.743	42.806	55.329
female_l7	31	13.839	4.506	4.111	11.343	17.198	23.959
male_l7	31	32.792	10.344	10.776	28.398	37.116	52.812
female_l8	30	34.64	9.081	13.919	29.998	38.534	55.27
male_l8	30	53.872	10.196	20.074	50.904	59.996	69.725
log_gdp	155	9.108	1.264	5.936	8.027	10.247	11.56
female_labour	41	69.378	16.66	19.86	59.493	82.045	90.08
male_labour	41	87.935	10.544	55.409	84.449	93.637	99.57

(i.e., $y = x$) is added to each sub-plot to assist us better understand the ratio on the share of female to the male – points above the straight line indicate that the country has more female who have at least a bachelor degree than male, vice versa. In the figure, the age group 25~34 has least number of points below this straight line – only two countries, Japan and Turkey, with Turkey is quite close to the straight line. On the other hand, the age group 55~64 has the most points. In the meantime, points in age group 55~64 centralizes on the right bottom corner of the whole figure, but the points in the age group 25~34 are scatter across the line.

The next two figures, Figure 5 and Figure 6 are five scatter plots for the proportion of people who major in STEM fields of the female to the male on different education levels. Figure 5 plots the combination of all education levels which are equal to or higher than Tertiary degree. Figure 6 contains four subplots which are for Tertiary degree, Bachelor degree, Master Degree, and Doctor degree, respectively. Again, we have a straight line $y = x$ to assist us understanding the ratio between the female and the male. Among all five figures, we can clearly see there is no point (i.e., country) that is above the straight line, which indicates in any country for any education level, there are always more male than female. Fortuantely, if we compare four subplots in Figure 6, we can see this difference can be relieved as the education level becomes higher.

Figure 7 talks about a different thing. It focus on the labour force participation. We plot the labour force participation of female and male, respectively, over the GPD per capita. The female subplot indicates a general increasing trend of labour force participation as the GDP increases. On the contrary, as the GDP increases, male labour force participation shows a first increasing than decreasing trend. However, if we focus on the range of y-axis, we can see that even if female's labour force participation keeps increasing, the male's labour force participation still enjoys a higher range. More explanantions of this figure is included in Appendix A.1.

4 Discussion

In this section, we will include interpretations on the results we provided in Section 3. This discussion connects the figures plotted in Section 3 with some existing literature to introduce my thoughts on the current

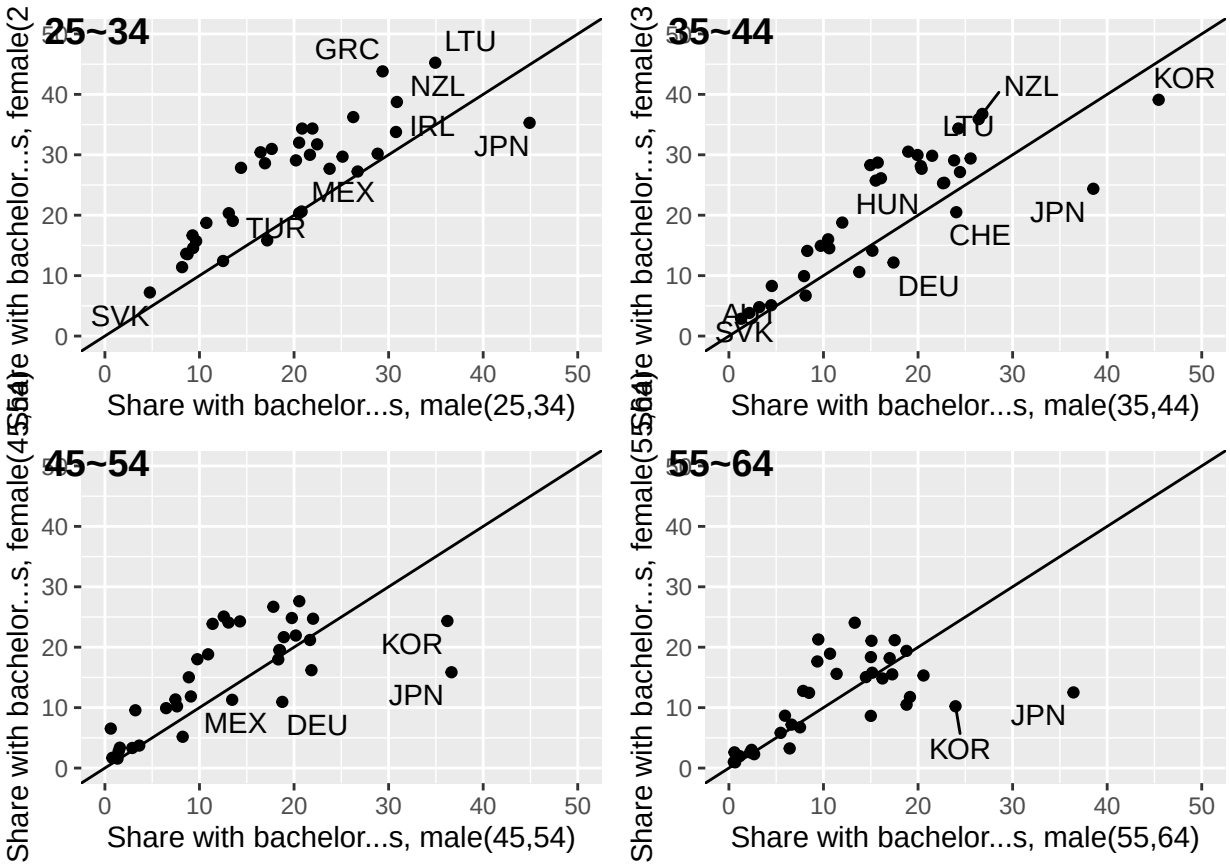

Figure 4: Shares of Men and Women with at Least a Bachelor's Degree in different age groups

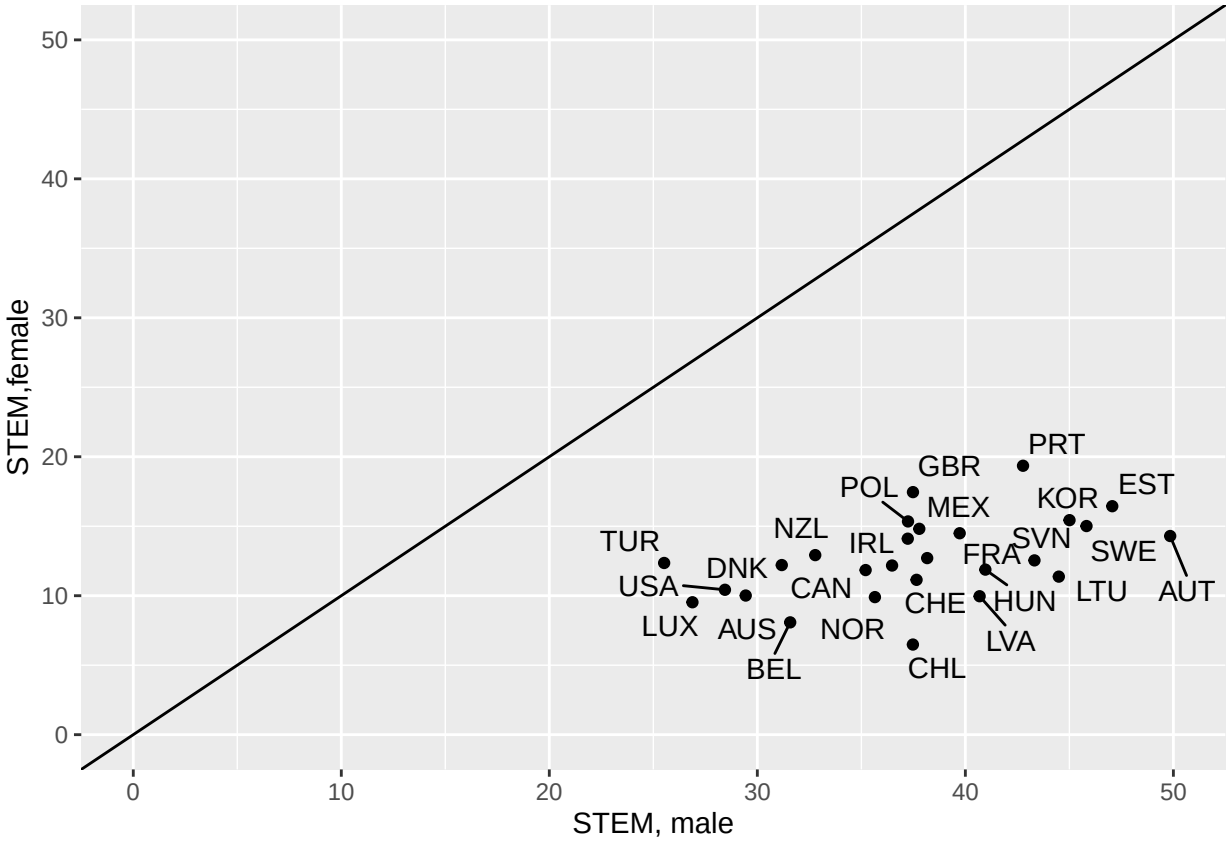

Figure 5: Gender Differences in STEM Education (Tertiary Degrees and above)

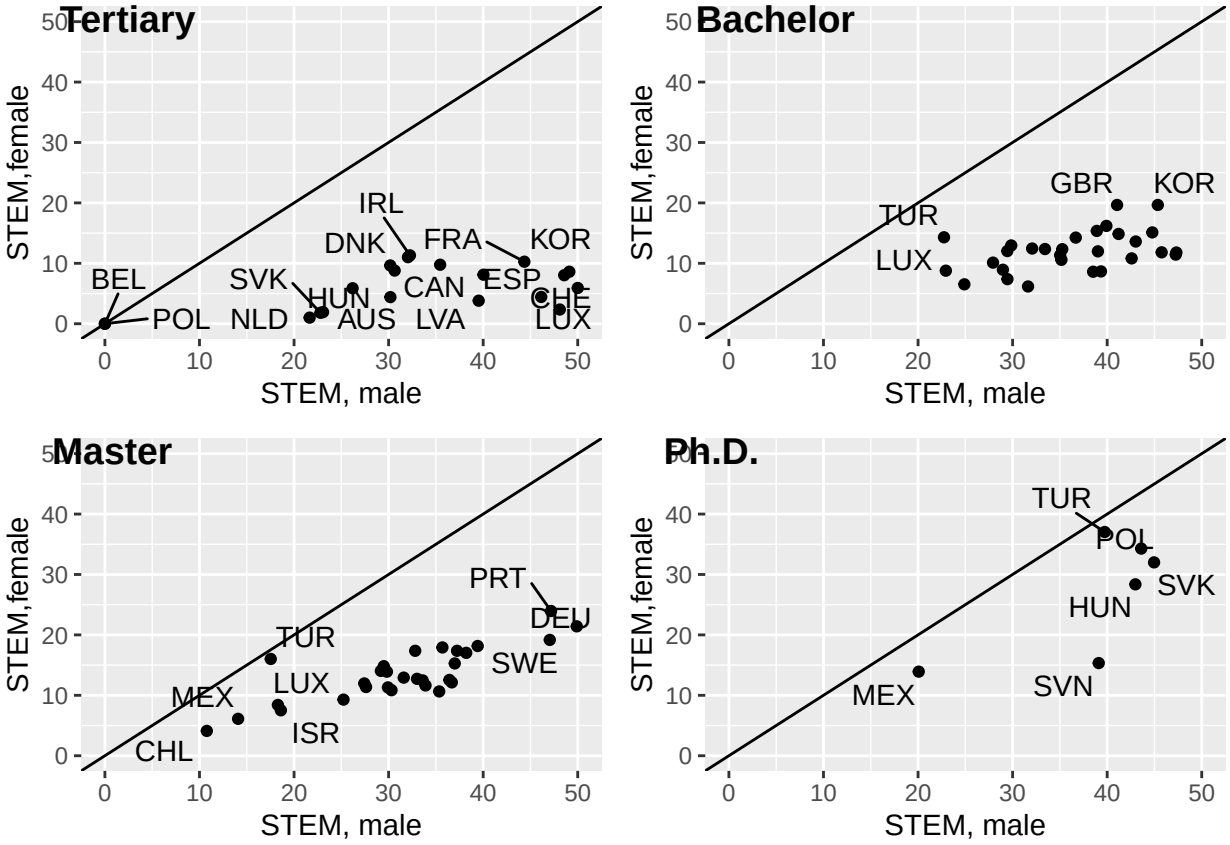

Figure 6: Gender Differences in STEM Education in different education level

Figure 7: Female and male labor force participation around the world

gender equality progress. According to what we have done in Section 3, we seperate the discussions into two parts, the education, and the job market.

4.1 Education: a distinct improvement

For the education, we can obviously observe the distinct improvements for gender equality, and even better. By Figure 4, we can see that as the age group goes down (i.e., younger age group), larger and larger share of female has attain a Bachelor or higher degree. Although this can be a result of the development for the higher education, since for male, the share is also growing, less and less points (i.e., countries) appear below the diagonal line can indicate the faster increasing share of the female. Meanwhile, according to Figure 5 and Figure 6, we can see that even though for STEM majors, whose graduates are supposed to find a job with higher salary, there are less female than male, as the education level increases, the difference on the share of female and male becomes smaller and smaller. Hence, we can claim that the gender equality has made a big difference in the education field.

4.2 Job market: still a long way to go

Things are quite different when it comes to the job market. Firstly, we can still look at Figure 5 and 6, even though in the higher education, the difference between female and male who majors in STEM becomes small, the Tertiary degree and the Bachelor degree still yields a quite large difference – male's share is almost three to four times of the female's. Unfortunately, these people compose of the major population. Also, even though for the Doctor degree, the difference sharply decreases, male's share is still around two times of the female's. Also, among all STEM majors, female systematically tend to choose non-mathematical or less-mathematical related majors, which results in a huge compact on the future salaries (Ceci et al. 2014). Thus, this observation strongly indicate the gender inequality in the job market.

Secondly, from Figure 7, we can see that even though as the GDP increases, the female labour force participation keeps increasing, if we compare the average of labour force participation, we can see a big difference between the male's participation and the female's. This observations is also supported in the literature. There is one interesting Figure from Bertrand (2020), Figure 4, as shown in Figure 8. This figure shows that if a job is scarce, what the share of male of female that agree with the statement that man should have more rights than a woman to the job. Again, a diagonal line to assist the comparison between female's opinion and male's opinion, we can still see that the share is competitive, which indicates the hidden rules of the job market – gender inequality is nature.

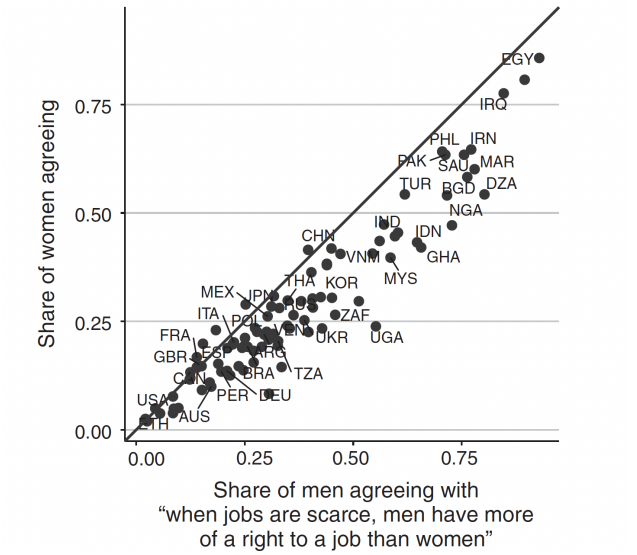

Figure 8: Shared beliefs about gender roles in the job market

4.3 Some potential reasons

We see from our analysis results that gender equality makes significant improvements in the education in the past years, but still a long way to go in the job market. But is this the case? There are some explanations in the literature. The first possible reason is the motherhood penalty. Researchers have conducted observational studies from different aspects for this problem. Kleven et al. (2019) states that when the first child is born in a family, the father will not be influenced even a little bit, but there will be a sharp decrease in average earnings of the father. Lundborg, Plug, and Rasmussen (2017) compares women with and without child(ren) to see their earnings difference by a VIF study in Denmark and they find that there is a significant different. In short, when a female becomes a mother, the job market will evaluate her in another set of rules and automatically decreases her competitiveness.

The second is the gender stereotypes. Bertrand (2020) states a surprising result that the job market itself does not discriminate female, and even is in favour of female – female do show power in female-dominated field, thought not statistically significant, but male do not show such things. So, where does the "discrimination" comes from? It might be the stereotype put on female.

In short, to solve the gender inequality shown in the job market, we should think from the potential reasons caused this inequality. Two possible reasons are widely discussed in the literature, the motherhood penalty and the stereotype put on female. More policies and emphasis should be made based on these two directions to further improve gender equality in the job market.

4.4 Ethics and bias

One of the major ethic and bias comes from the scope of the data collection. As introduced in the paper, data are collected from OECD member countries, which only contains 38 countries in total. Unfortunately, China and Russian, two biggest countries in the world are also excluded, which means the scope of the data set is too small to be representative. In other words, the conclusions drawn from the figures, are suitable for the OECD member countries only and cannot be generalized to the whole world.

Another thing is that the time coverage of the data set we utilized is quite limited – it only occupies 2016 and 2017. The analysis should include more data from other years, especially recent years. In this way, we can clearly see the trend of changes and we can also observe the influence caused by COVID-19.

4.5 Weaknesses and next steps

The weakness of this paper is the coverage of the job market elements. For the education, we have include a detailed discussion from different age groups, different education levels, and different majors to compare the difference between female and male. However, for the job market, even though we have include a comprehensive discussion from the literature, we do not use the data to strongly support our argument.

Hence, in the future, we may want to collect some data such as salaries, the proportion of employers from some big names, the female occupation of the leader positions in compnaies, etc. to discuss on the gender equality in the job market. In the meantime, we may also want to address the ethic and bias issues discussed above.

A Appendix

A.1 Replication

Three figures in this paper is a replication work of the original paper Bertrand (2020).

- The left top subplot of Figure 4 is a replication work of Figure 2 in Bertrand (2020).
- Figure 5 is a replication work of Figure 3 in Bertrand (2020)
- Figure 7 is a replication work of Figure 1 in Bertrand (2020)

The first two figures do not have any problems. However, when plotting the last figure, I found the data provided by the author were not in accordance with the data they used to draw this figure. In the meantime, the code they provided were not reproducible since some intermediate files are missed. Based on the existing files they provided, I am 100% confidence I did not use the wrong data set because I double-checked multiple times, went through their `stata` codes to make sure everything was fine. I also tried to reproduce other figures in their paper, but I found that they even did not provide related data – even though they put a folder there. I tried to contact the author but did not get any response. Thus, I would really appreciate it if you could understand this condition and do not take points off for this embarrassing conditions.

By the way, since the figure is supposed to be a comparison between two years, 1990 and 2017, the legend shows two different year with different point shapes and colours. However, due the missing values, the scatter plot itself only contains one year. This is not a mistake, but plotted intentionally.

References

Bank, World. 2022. "World Development Indicators." *WDI - Home*. http://datatopics.worldbank.org/world-development-indicators/.

Bertrand, Marianne. 2020. "Gender in the Twenty-First Century." In *AEA Papers and Proceedings*, 110:1–24.

Ceci, Stephen J, Donna K Ginther, Shulamit Kahn, and Wendy M Williams. 2014. "Women in Academic Science: A Changing Landscape." *Psychological Science in the Public Interest* 15 (3): 75–141.

Huntington-Klein, Nick. 2021. *Vtable: Variable Table for Variable Documentation*. https://CRAN.R-project.org/package=vtable.

ILO. 2022. "ILOSTAT Data Tools to Find and Download Labour Statistics." *ILOSTAT*. https://ilostat.ilo.org/data/.

Kassambara, Alboukadel. 2020. *Ggpubr: 'Ggplot2' Based Publication Ready Plots*. https://CRAN.R-project.org/package=ggpubr.

Kleven, Henrik, Camille Landais, Johanna Posch, Andreas Steinhauer, and Josef Zweimuller. 2019. "Child Penalties Across Countries: Evidence and Explanations." In *AEA Papers and Proceedings*, 109:122–26.

Lundborg, Petter, Erik Plug, and Astrid Würtz Rasmussen. 2017. "Can Women Have Children and a Career? IV Evidence from IVF Treatments." *American Economic Review* 107 (6): 1611–37.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

OECD. 2022. "Gender Equality." *OECD*. https://www.oecd.org/gender/.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

RStudio Team. 2020. *RStudio: Integrated Development Environment for r*. Boston, MA: RStudio, PBC. http://www.rstudio.com/.

Slowikowski, Kamil. 2021. *Ggrepel: Automatically Position Non-Overlapping Text Labels with 'Ggplot2'*. https://CRAN.R-project.org/package=ggrepel.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Evan Miller, and Danny Smith. 2022. *Haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files*.