Replicate and extend the results of the article "Relationship of gender differences in preferences to economic development and gender equality"

Sara Cerioli   Andrey Formozov

Abstract

This study replicates and extends the article of Falk and Hermle published in *Science* (2018), where the gender differences in economic preferences and their relation to economic development and to gender equality of the countries were measured. In the original paper, the authors use the data from the Gallup World Poll 2012, which included a Global Preference Survey conducted on almost 80000 people in 76 countries all around the world. The dataset covers almost 90% of the world population representation, with each country having around 1000 participants answering questions related to their time preference (patience), altruism, will of risk taking, negative and positive reciprocity, and trust. The full dataset is available in its integrity only with a license to be paid. The free version of the data is available not in the rawest form possible and has a smaller number of predictors. Nevertheless, we managed to closely reproduce the original findings and conduct a further investigation of the methodology and the results obtained by the authors in the original study. Moreover, using a different indicator for the gender equality of the countries, the correlation between gender differences and gender equality of the country disappears, when a conditional analysis on the economic development is performed.

1. Introduction

Gender differences in the economic behaviors, such as happiness (Schneider *et al.* 2012), competition (Croson & Gneezy 2009; Gneezy *et al.* 2009; Klonner *et al.* 2021), or work preferences (Beblo & Görges 2018) have been studied in sociology, psychology and economics for many decades. Any scientific knowledge in this regard and its relation to gender inequality is used nowadays as arguments and counter-arguments for decision and policy-making. Gender inequality topics are becoming an integral part of the agenda for many public and private institutions and organizations, and it is essential for the stakeholders to reveal, estimate, monitor, and prevent gender inequalities on individual, group and country levels.

The study of gender differences on a world scale per se is challenging. One of the challenges that hampers the progress is the lack of large and heterogeneous datasets across different social groups and countries. The Gallup World Poll 2012 included a Global Preference Survey conducted on almost 80000 people in 76 countries all around the world that aimed to fill this gap: Covering nearly 90% of the world population representation, with each country having around 1000 participants answering questions related to their time preference (patience), altruism, will of risk-taking, negative and positive reciprocity, and trust. The dataset provides a unique insight into the economic preferences of a heterogeneous number of people.

The article that conducted the analysis of this dataset was published in the Quarterly Journal of Economics (Falk *et al.* 2018). It focused on general questions about the economic preferences distributions in different countries, exploring several covariates from the Gallup World Poll. The subsequent article (Falk & Hermle 2018a), replicated in this work, focused explicitly on the gender differences arising from the previous study and reported the evidence for the relationship between gender differences in economic preferences, economic development, and gender equality across many countries. The authors propose two competing hypotheses to be tested: The first one is that the gender differences will decrease for more economically developed and gender-equal countries, because social roles related to gender are attenuated. The second hypothesis is that, on the contrary, for more economically developed and gender-equal countries, the gender differences will increase because the gender-neutral goal of subsistence is removed and thus people can pursue their real preferences. Their analysis shows a positive correlation of gender differences in preferences with economic development (expressed as Log GDP p/c) and with the gender equality of the countries, thus favoring the hypothesis that predicts the increase of the differences as women and men obtain sufficient access to the resources to develop and express their intrinsic preferences independently. The authors composed a joint measure of gender equality of different countries using two officially recognized indexes to measure gender equality (WEF Global Gender Gap and UNDP Gender Inequality Index), together with two other less complete indicators, providing little to no discussion about the validity and robustness of their approach, and without establishing a link to the existing literature with this regard. Besides, the power and limitations of the officially recognized indicators themselves is an open scientific question, and, even being the best estimates available on global scale, these are far from being a complete representation of gender inequality (Anand & Sen 1995; Permanyer 2013; Klasen 2017). The present work is dedicated to close this methodological flaw and rise awareness about possible implications on the results, and drawing conclusions regarding the relationship between gender differences in economic preference and gender equality in the original study.

In this study, we first analyse the gender equality index that was used by the original authors and the sub-indexes involved into the study, highlighting its problematic. Second, we conduct a pure replication of the article in R-language. Third, we extended the original article using a robust linear regression to address the non-normality of the data. Fourth, we demonstrate that the fundamental finding is not stable when trying to replace the gender equality index introduced by the authors and discuss [last part will be improved after completion of the results section and dicussion].

2. Results

A composed Gender Equality Index and related problematics

The authors in the original article composed a joint measure of gender equality that they denoted as Gender Equality Index (GEI). To reveal its actual structure, we visualized its composition using a diagram shown in Figure 1. Four indicators were used to compose this joint measure by employing Principle Component Analysis and taking the first component as a proxy of gender equality. Two of them are indexes officially approved by international organizations, the Global Gender Gap from the World Economic Forum Global Gender Gap Report 2015, and the Gender Inequality Index from the UNDP Human Development Report 2015; one widely used measurable, the ratio of female and male labor force participation, taken from the World Bank database; and lastly, a measurable newly constructed by the authors, the time since women suffrage, taken from the Inter-Parliamentary Union Website, presumably to track long-term influences of the guaranteed right to vote as a proxy of gender equality. The WEF Global Gender Gap has a total of fourteen sub-indexes, grouped and weighted into four categories: economic participation and opportunity, political empowerment, educational attainment, and health and survival. The UNDP Gender Inequality Index follows a similar logic to cover the same categories describing different aspects of the human life, but using only five sub-indexes in total, that are two for health and reproduction-related issues, and three others for the remaining categories.

The UNDP Gender Inequality Index has been criticized by several authors (Anand & Sen 1995; Permanyer 2013; Klasen 2017). This index is said to be very highly related to the economic development, as it includes reproductive health indicators that can penalize less-developed countries, and it has a "inexistent" measure of welfare loss of inequality, because it is based on a calculated gender equality measure that is reported
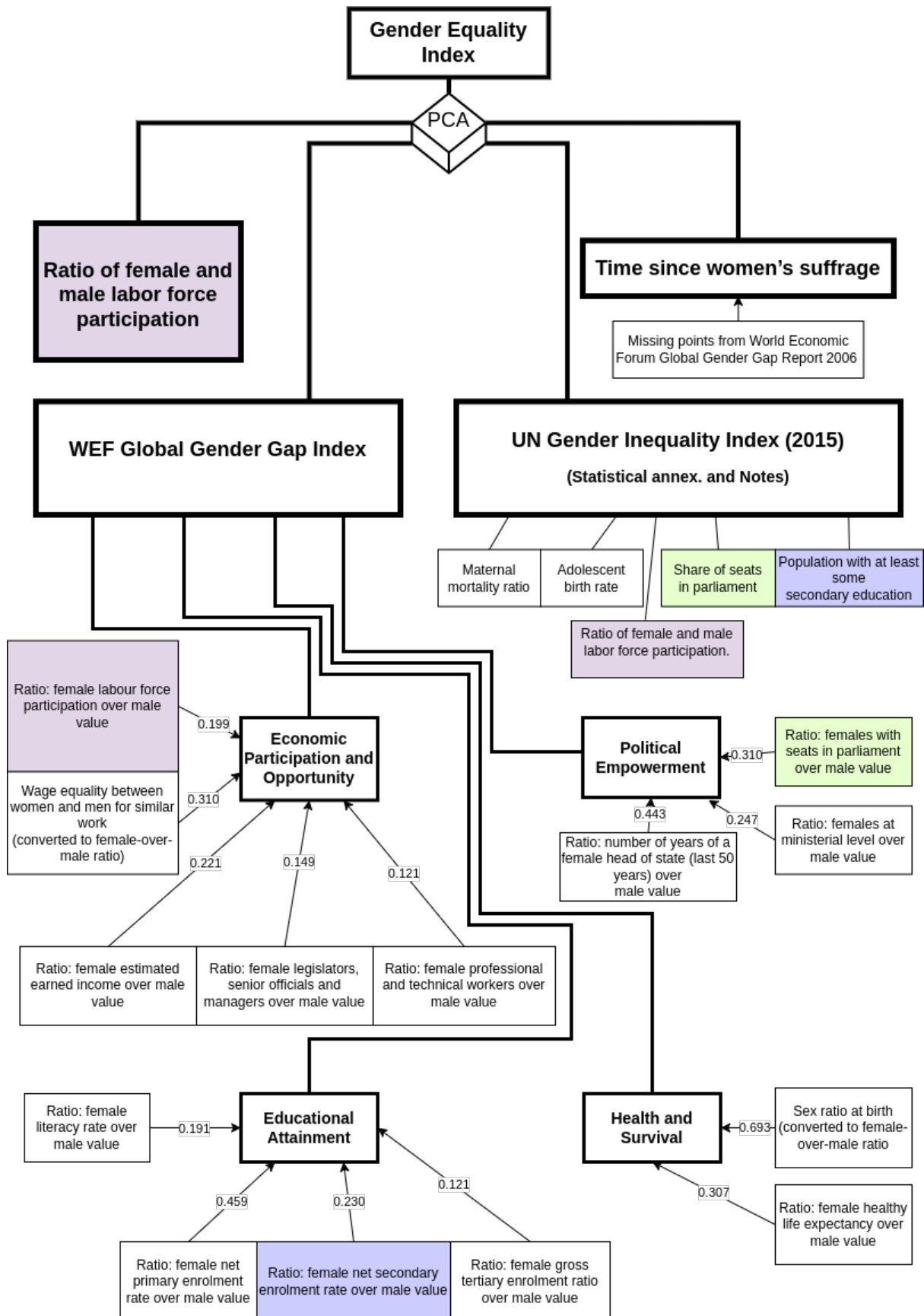
Figure 1: Gender Equality Index decomposed in its sub-indexes.

nowhere. One of the coefficients that has a highest priority (~30% of the sub-index) in calculation of the economic participation and opportunity, called "wage equality between men and women for similar work" – indicating how much women receive with respect to men for the same work – is based on a best experts' guess, not corrected on the difference in occupation (Halkos *et al.* 2021).

The power and limitation of WEF Global Gender Gap index remains mostly undiscussed in the academic literature, with a few exceptions (Barns & Preston 2010; Piper 2019). The index is thought to be the least dependent to the economic development of a country, since it measures the gap between male and female access to resources and opportunities ('The global gender gap report' 2006).

Furthermore, the "time since women suffrage" index was introduced by authors to track long-lasting effects of the right to vote. This is based on the assumption that during the course of the time development has always a positive effect and its magnitude is proportional to the time since women suffrage was established, and the magnitude of this effect is proportional to the time since women suffrage. It can be argued that, even after the right to vote, many discriminating laws were still in presence, and the alignment of law together with executive branch of the government and elimination of discrimination takes more time – for example, despite to the right to vote, the right to work can be suppressed for several decades. This assumption that suffrage played a long lasting effect on the balance in gender equality sounds reasonable but requires further investigation to be used as a robust estimator. Worthy to mention also the question about suffrage and race, that for many countries is strongly connected (Carruthers & Wanamaker 2015; Yang 2020).

The question why such joint Gender Equality Index is better or at least not worse than the widely used indexes such as WEF and UNDP GII was unanswered in the original article. As one can see in Figure. 1, the components of the joint Gender Equality Index used in the original study contains repetitions. The two indexes from WEF and UNDP share three sub-indexes, here indicated with different colors: ratio of female and male labor force participation (yellow), share of seats in parliament (green) and enrollment into secondary education (blue). As a third variable to construct a joint Gender Equality Index, the authors used the ratio of female and male labor force participation, already included into the previous two indexes. It is true that PCA technique permits to aggregate the indexes even in presence of high correlations between them [REF –> *Isn't it obvious from the techniques itself?*]. However, in the present case, such a procedure leads to the imbalance and prioritization of these specific repetitive indexes (especially female and male labor force participation) over other factors, that were already balanced in the design of WEF and UNDP GII indexes.

Based on the identified methodological flaws and absence of validation and justification, we doubt the possibility to use Gender Equality Index as a reliable proxy for gender equality. Therefore, we will use GEI only to match the replication with the original analysis. The new analysis will use only established WEF GGG Index and the UNDP GII, in their relationship to the gender differences in economic preferences and the economic development.

Along with WEF GGG Index and UNDP GGI, other indexes exist, such as the Gender Development Index, that was discussed in Klasen (2017) and recently introduced in the UNDP report. It captures three dimensions in terms of health, knowledge and living standards, separately for male and female (as it is defined as the ratio of the Human Development Index for female divided by the male one). The life expectancy, expected years of schooling and mean years of schooling, and GNI per capita are calculated within these dimensions. We include this index in the analysis in the next section too.

Pure replication and extended analysis

Before showing the results of the extended analysis, we briefly describe the methodology of the original article and provide information on the pure replication, with the identification of found discrepancies.

To conduct the replication, we requested the Gallup World Poll data. The full dataset is under restricted access: education level and household income quintile on the individual level of the participants is not available in the open access version (for more information see Appendix, Section 2, "Data Collection, Cleaning, and Standardization"). As the (Falk & Hermle 2018a) article addresses the gender differences, the main focus is on that individual variables provided in the dataset (education level, income quintile, age, and subjective math skills) are taken as control variables, meaning that the presence of these may not affect the result of the

correlation (Falk & Hermle 2018b). Therefore in the replication study, for each country, a linear model is created using the equation:

$$p_i = \beta_1^c female_i + \beta_2^c age_i + \beta_3^c age_i^2 + \beta_4^c subjective\ math\ skills_i + \epsilon_i$$

This results in 6 models – one for each preference measure, p_i – having intercept and 4 coefficients, each of them being related to the variable in the formula above. The coefficient for the dummy variable "female", β_1^c, is used as a measure of the country-level gender difference. Therefore, in total, we have 6 coefficients that represent the preference differences related to the gender for 76 countries. To summarise the gender differences among the six economic preferences, a principal component analysis (PCA) is performed on the gender coefficients from the linear models. The PCA is a dimensionality-reduction technique which allows to "reshape" the 6 coefficients into other mixed components that maximizes the variance. The first component of the PCA has then been used as a summary index of average gender differences in preferences, in a similar fashion as for the Gender Equality Index described in the previous paragraph.

In addition to this, we questioned the normality of the input data and therefore of the robustness of the linear regression performed above. One way to address the problem of the long tails resulting in non-normally distributed residuals can be to choose a model with less restrictive assumptions, for instance the robust linear regression. Thus, in the extended analysis, the models for each countries were created by using the robust linear regression, and same substitution happens for every model created using linear regression.

The correlation between economic development expressed in Log GDP and gender equality indexes in the countries is very strong (see Fig. 2) and one can be used as a rather good estimator of another. Because of this fact, the presence of strong correlation between gender differences in preference(s) and gender equality indexes is expected if only two variables are considered in the regression. To study the effect of both GDP p/c and gender equality at the same time, one should incorporate all three variables into the multiple linear regression with both factors as explanatory variables, like:

$$avgGenderDiff_{country} = \alpha + \beta_1\ logGDPpc_{country} + \beta_2\ genderEquality_{country}$$

Alternatively, to separate the contribution of the gender differences regressed on economic development, one can perform a regression conditioning on the gender quality indicators (WEF GGG Index, UNDP GII and Gender Development Index), and then conditioning on economic development (Log GDP p/c). The theorem from Frisch–Waugh–Lovell (Frisch & Waugh 1933; Lovell 1963) guarantees that the coefficients found from this conditional analysis are the same as the ones found for a regression of the gender differences on both economic development and gender equality index of the countries.
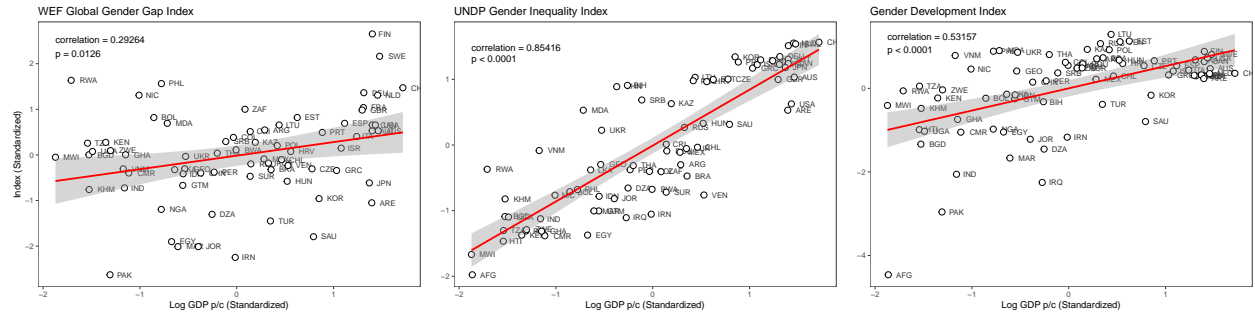

Figure 2: Correlation between gender equality indicators and economic development by country.

In Table 1, we compare the results obtained from our replication analysis (meaning, using the OLS method as in the original article) and the extended analysis (substituting the OLS with the robust linear regression) to the results found in Falk & Hermle (2018a), Fig. 2A-F.

Table 1: Comparison of the conditional analysis results for original, replicated and extended study . Reported are the **slopes** of the linear regressions and the corresponding p-value. Significance levels: ≤ 0.001 (***), ≤ 0.01 (**), ≤ 0.05 (*).

Variable	Residualized on	Original	Replication	Extended
Log GDP p/c	Gender Equality Index	0.5258***	0.5003***	0.4862***
Gender Equality Index	Log GDP p/c	0.3192**	0.3358***	0.3432**
WEF Global Gender Gap	Log GDP p/c	0.2327**	0.2234*	0.2106*
UN Gender Equality Index	Log GDP p/c	0.2911	0.3180	0.3017
F/M in Labor Force Participation	Log GDP p/c	0.2453*	0.2206*	0.2034*
Years since Women Suffrage	Log GDP p/c	0.2988**	0.1879*	0.1929*

Most of the time, we find very similar slope coefficient and statistical significance between the variables of our study (both replicated and extended) and the results from the original study, with only two exceptions: One is related to the indicator "WEF Global Gender Gap", and the other to the "Years since women suffrage". Regarding the last one, the differences can be explained by the imputation of the data used. The dataset from which this indicator is coming from, as a matter of fact, can be interpreted non unanimously (see the paragraph "Gender Equality Index" in the Appendix and discussion sections) and we can't draw any conclusion from it. The difference with the WEF Global Gender Gap is instead pretty mysterious, since the data were taken from the same source and no substantial change is supposed to affect it. *NOTE: I still have no clue how to conclude this sentence properly.*

Reassured by the similarity of the results of our replication, we proceed to explore the correlation of the summarised gender differences to the economic development, when conditioned on the gender equality indicators indicated in the paragraph before as adequate (WEF GGG Index, UNDP GII, and GDI), the correlation of the summarised gender differences to the single gender indicators, when conditioned on economic development, and then the same but for single preferences.

In Fig. 3, we summarise the results of the conditional analysis for the three main indicators that were further analysed, that are WEF GGG Index, UNDP GII and GDI. The gender differences are strongly and statistically significantly correlated with the economic development when the conditional analysis is performed on the single gender equality indicators, thus indicating that the economic development of a country seems to play a key role in the measured gender differences in the economic preferences. On the contrary, the correlation of the gender differences to the gender equality indexes of the country, conditioned on the economic development is only statistically significant for WEF GGG Index conditioned on Log GDP p/c (p-value = 0.0241), while for UNDP GII and GDI the correlation is not statistically significant.

To investigate the role of economic development and gender equality on single preferences, we performed the same analysis but without PCA-aggregating preferences into a single measure. As one can see in Table 2, alongside with the results of aggregated gender differences, the economic development still has in most cases a strong correlation to the single preferences when conditioned on the gender equality of the country. The only exception is the UNDP GII indicator, that showed a very strong correlation to the economic development. *How to conclude this? It is an exception and we can't ignore it*

Table 2: **Log GDP p/c conditional on the gender equality indicators**, regressed on each single economic preference gender difference coefficient. Reported are the correlation terms and their significance levels: ≤ 0.001 (***), ≤ 0.01 (**), ≤ 0.05 (*).

Variable	WEF (Extended)	UNDP (Extended)	GDI (Extended)
Trust	0.5174***	0.1930	0.5045***
Altruism	0.5255***	0.3527***	0.4477***
Positive Reciprocity	0.2898*	0.2054	0.1870

Variable	WEF (Extended)	UNDP (Extended)	GDI (Extended)
Negative Reciprocity	0.3974**	0.0706	0.3858***
Risk Taking	0.3469**	0.0262	0.3199**
Patience	0.3742**	0.1046	0.3337**

Lastly, Table 3 summarizes the results of the regression of the gender differences in single economic preferences with relation to gender equality indexes conditioned on economic development. No statistically significant correlations were observed for all gender equality indexes and preferences except the single case of WEF index -Altruism combination and of the UNDP index -Risk Taking, when controlling for the economic development. Thus, no reliable association between the gender differences in economic preferences and gender equality indexes is found.

Table 3: **Gender equality indicators, conditional on Log GDP p/c**, regressed on each single economic preference gender difference coefficient. Reported are the correlation terms and their significance levels: ≤ 0.001 (***), ≤ 0.01 (**), ≤ 0.05 (*).

Variable	WEF (Extended)	UNDP (Extended)	GDI (Extended)
Trust	0.1325	0.2160	0.0794
Altruism	0.3561**	-0.0421	0.1021
Positive Reciprocity	0.0396	-0.0402	0.1978
Negative Reciprocity	0.1680	0.1742	-0.1033
Risk Taking	0.0349	0.2192*	-0.0261
Patience	0.2312	0.1705	0.0812

3. Discussion and conclusions

The original study indicates that higher economic development and higher gender equality are associated with an increase in the gender differences in preferences, and therefore rules out the social-role theory over the post-materialistic one: When more resources are available to both men and women, the expression of the gender specific preferences can be seen.

Our analysis focuses on two main points: First, we aimed to replicate the results of the original article, in order to set the methodology and prove that the datasets, even if not being the same as the ones used by the original authors, can lead to similar results. Second, we extended the analysis by introducing a new gender equality indicator and by performing the analysis using the robust linear regression. While the use of the new dataset is due to our critics to the construction of the original authors Gender Equality Index, supported by a vast literature on an ideal gender equality indicator (Anand & Sen 1995; Barns & Preston 2010; Permanyer 2013; Klasen 2017; Piper 2019; Klonner *et al.* 2021), the robust linear regression was supported by the diagnostic made on the data fitted by an OLS and the fact that this gave signs of non-normality in the data that could potentially influence in a much stronger way than wished the results of the linear fit.

Starting from the analysis of the Gender Equality Index built by the authors, we reveal some inconsistencies with the methods to build such indicators: The preference of simplicity over complexity to help policy-makers understand the indicator, and the avoidance of inclusion of too many - or too strong – sub-indexes that relates directly to the economic development of the country (Anand & Sen 1995; Permanyer 2013; Klasen 2017; Klonner *et al.* 2021). Thus, we decided to focus only on the most complete indicators that were already available in the original article, meaning the WEF Global Gender Gap Index and UNDP Gender Inequality Index, adding a new indicator following the suggestions of Klasen (2017), the Gender Development Index.

Moreover, since a strong correlation has been found between each of the indicators used (including the Gender Equality Index built by the original authors), we focused on the gender differences regressed on economic
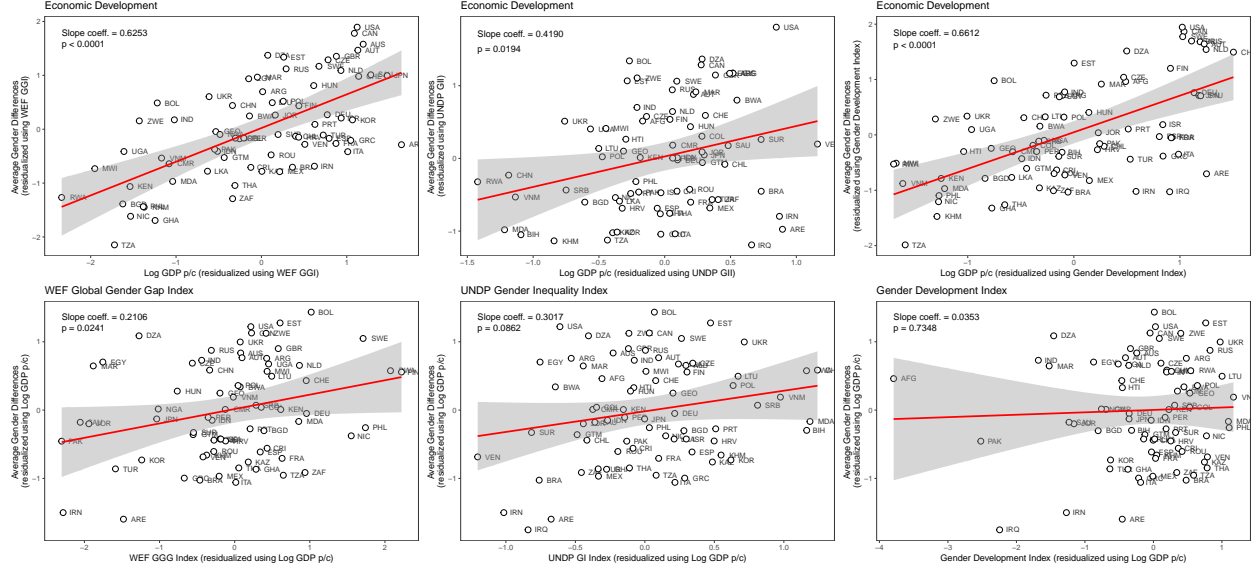
Figure 3: On the top line, gender differences are regressed on the economic development conditioned on gender equality, for the different indicators (WEF GGG Index, UNDP GGI, and Gender Development Index). On the bottom line, the corresponding values of gender differences regressed on the gender equality indicator conditioned on economic development.

development conditioned on each of the gender equality indicators, and vice-versa, to extract the correlation of the differences depurated of the confounding terms of economic development and gender equality of the countries.

We prove that, using a different indicator for the gender inequality in different countries, the results found by Falk & Hermle (2018a) are less straightforward. To summarise the main findings on this regard:

- The summarised gender differences shows a positive, strong and statistically significant dependence to the economic development of a country, when conditioned on the corresponding gender equality;

- The summarised gender differences shows a positive weak correlation to the gender equality of a country, when conditioned on their economic development. The result is not statistically significant for UNDP GII and GDI, and it is significant at the 5% level (p-value = 0.0241) for the WEF GGG Index.

- The results for the single preferences regressed on economic development, conditioned on gender equality, are far from being clear: All the preferences, except for "altruism", show different behavior and different statistical significance, while "positive reciprocity" seems to be the variable with the least correlation, according to all the indicators. The results produced conditioning on WEF and on GDI seem to be more in agreement among themselves. To be noticed, however, the very strong correlation of UNDP GII and economic development, that might be the cause for such a weak association of the preferences to the economic development, when regressed on it.

- The results for the single preferences regressed on the gender equality, conditioned on the economic development, are all in agreement with each others, displaying weak to no correlation, and no statistical significance. The only two exceptions are for the preference "altruism" that shows a positive, moderate and statistically significant (p-value = 0.006) correlation with the WEF GGG Index, in disagreement with the UNDP GII and the GDI; and the preference "risk taking" showing a positive, weak, and statistically significant at 5% level correlation with the UNDP GII, also this in disagreement with the other two indixes.

Looking at these results, we can deduce that, for countries having the same gender development, the gender differences are depending from the economic development of the country, meaning that richer countries have

8

more gender differences than poorer countries, gender development being equal. On the other hand, for countries having the same economic development, the gender differences are independent from the gender development of the country, meaning that there are no gender differences arising from countries having same economics but different gender development.

From this, one could therefore assume that the economic development is the country-level indicator associated with higher changes in gender differences, rather than the gender development of a country. From such a simple analysis is therefore not possible to extract any information regarding the reason for higher gender differences in more economically developed countries. We can only see that, for countries with similar economics, those differences don't exist regardless of how gender developed that countries are. The reason behind the differences might be related to purely economical conditions, that would also make sense since the gender differences here studied are "economical preferences". Can it be that the differences arise where the economic is more developed because of marketing reasons? After all, all the economics here analysed are based on capitalistic systems, meaning that the bigger the economic, the bigger the market. It is an interesting hypothesis that would require further analysis.

References

Anand, S. & Sen, A. (1995). Gender inequality in human development: Theories and measurement. *Occasional Papers 19, Human Development Report Office, New York.*

Barns, A. & Preston, A. (2010). Is australia really a world leader in closing the gender gap? *Feminist Economics*, **16**, 81–103. Retrieved from https://doi.org/10.1080/13545701.2010.530607

Beblo, M. & Görges, L. (2018). On the nature of nurture. The malleability of gender differences in work preferences. *Journal of Economic Behavior & Organization*, **151**, 19–41.

Carruthers, C.K. & Wanamaker, M.H. (2015). *Municipal housekeeping: The impact of women's suffrage on public education.* National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w20864

Croson, R. & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, **47**, 448–74.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D. & Sunde, U. (2018). Global evidence on economic preferences. *The Quarterly Journal of Economics*, **133**, 1645–1692.

Falk, A. & Hermle, J. (2018a). Relationship of gender differences in preferences to economic development and gender equality. *Science*, **362**.

Falk, A. & Hermle, J. (2018b). Supplementary materials of relationship of gender differences in preferences to economic development and gender equality.

Frisch, R. & Waugh, F.V. (1933). Partial time regressions as compared with individual trends. *Econometrica*, **1**, 387–401. Retrieved from http://www.jstor.org/stable/1907330

Gneezy, U., Kenneth, L.L. & List, J.A. (2009). Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica*, **77**, 1637–1664.

Halkos, G., Alba, J.M. de & Todorov, V. (2021). Economies' inclusive and green industrial performance: An evidence based proposed index. *Journal of Cleaner Production*, **279**, 123516.

Klasen, S. (2017). UNDP's gender-related measures: Current problems and proposals for fixing them.

Klonner, S., Pal, S. & Schwieren, C. (2021). Equality of the sexes and gender differences in competitiveness: Experimental evidence from a traditional society with gender-balanced norms.

Lovell, M.C. (1963). Seasonal adjustment of economic time series and multiple regression analysis. *Journal of the American Statistical Association*, **58**, 993–1010.

Permanyer, I. (2013). Are undp indices appropriate to capture gender inequalities in europe? *Social Indicators Research*, **110**, 927–950.

Piper, A. (2019). An investigation into the reported closing of the nicaraguan gender gap. *Social Indicators Research*, **144**, 1391–1413.

Schneider, U., Pfarr, C., Schneider, B.S. & Ulrich, V. (2012). I feel good! Gender differences and reporting heterogeneity in self-assessed health. *European Journal of Health Economics*, 251–265.

The global gender gap report. (2006). Retrieved from http://www3.weforum.org/docs/WEF_GenderGap_Report_2006.pdf

Yang, M. (2020). An intimate dialog between race and gender at women's suffrage centennial. *Humanities and Social Sciences Communications*, **7**.