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 attempts to replicate and extend the results of the article of Falk & Hermle (2018a) measuring the gender differences economic preferences relating them to economic development and to gender equality of the countries. In the original paper, the authors use 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 dataset is available in its integrity only with a license to be paid. The free version has only partial data that can also be used for this purpose because, according to the FH study, the gender differences can be studied also only taking in consideration a smaller number of predictors (according to the supplementary material, see (Falk & Hermle 2018b)). In this replication study, therefore, we use only a subset of predictors that are made publicly available to check whether the results can still be reproduced and are consistent. The outcome of the replication is that we see similar results as the ones obtained by the original authors for the relationship of gender differences and the economic development, but with differences (some times minor, some times significantly large) regarding the gender equality, especially when comparing the results of the single indexes building the general Gender Equality Index. Beyond the replication, we have extended the analysis using a robust linear regression instead of the OLS used by the authors, finding that the results are not robust under this change of model. Moreover, using a different measurement for the gender equality of the countries, the Gender Development Index from the UN Human Development reports, the correlation between gender differences and Gender Development of the country disappears, when a conditional analysis on the Economic Development is performed.

1. Introduction

Gender differences are nowadays extensively used as arguments and counter-arguments for decision and policy making, and the differences concerning the economic behaviors, such as happiness (Schneider *et al.* 2012), competition (Croson & Gneezy 2009; Gneezy *et al.* 2009), or work preferences (Beblo & Görges 2018), are being studied in many sectors of the economy and economy-related fields.

One of the problems common for many experiments in social sciences is the lack of large and heterogeneous datasets that can be used to check for such differences reducing some of the bias induced, for example, by having students or specific sets of people interviewed for the study.

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 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 dataset provides a unique insight in the economic preferences of a heterogeneous amount of people. The original study published in the Quarterly Journal of Economics (Falk *et al.* 2018 p. 133 (4) pp. 1645-1692) focused on more general questions about the economic preferences distributions in different countries, trying to explore different covariates from the Gallup World Poll. While, the subsequent article, replicated in this work, focused specifically on the gender differences arising from the previous study.

The main question that the article wants to study is whether the gender differences in economic preferences increase or decrease as the economic development and gender equality of the countries increase. In the first scenario, the gender differences increase as the economic development increases because the gender-neutral goal of substistence is removed, and therefore the real preferences can be pursued. Moreover, since those countries are usually also the ones with more gender equal societies, we would have more women and men allowed to express their desires and preferences in an independent way. This would be the so-called resources hypothesis. On the other hand, there is the social role hypothesis stating that the more economically developed and gender-equal the country, the smaller the gender differences because of the attenuation of the social roles related to the genders. The conclusion of the article is that the trends in the data shows a positive correlation of gender differences with GDP p/c and with the gender equality of the countries, and thus "confirming" the resources hypothesis.

NOTE: Here I need to add something about our extended analysis

2. Methods

Overview

We replicate the results using the R programming language version 4.0.3 (2020-10-10), and its open-source IDE RStudio for an easy access of the code. In the Appendix, we include a list of the packages used and their corresponding versions.

For an in-depth look at the replication methods, including the data collection, cleaning, and standardization of the different indicators and data sources, and the data analysis details, please refer to the Supplementary Material.

Data Analysis

The article uses the following methods commonly accepted in the field:

- Linear regression for each Country for each preference to extract the gender coefficient as a measure of the gender differences.

- Principal Component Analysis on 6 gender coefficients to summarize an overall measure of the gender differences, and 4 gender equality indexes of the countries to summarize an overall Gender Equality Index.

- Variable Conditioning to separate further between economic development and gender equality in the country.

The main focus of the extended analysis that we performed was on the model used to extract the gender coefficients, and on the different index for the gender (in)equality.

Linear Models Diagnostic and the Robust Linear Regression

As already mentioned in the previous paragraph, part of the data to reproduce the article is under restricted access: education level and household income quintile on the individual level are not available in open access version. As the (Falk & Hermle 2018a) article addresses the gender differences, the main focus is on that individual variable and all the others provided in the dataset (education level, income quitile, age, and

subjective math skills) are taken as control variables, meaning that the presence of these variables may not affect the result of the correlation.

The linear model for each country 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 weights, each of the weight being related to the variable in the formula above. The weight 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 weights that represent the preference difference related to the gender for 76 countries.

We now question the normality of the input data and therefore of the robustness of the linear regression performed as described. In order to verify our doubts, we used the diagnostic plots for the linear regression, in particular focusing on the so called "Normal Q-Q" plot. This plot helps determining the normality of the residuals by looking at their distribution along a straight diagonal line.

In the Figures "normQQ" we present some example of these plots, chosen randomly for some countries and some preferences, where one can deduce that the normality of the residuals is not respected.

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. In the robust linear regression, each datapoint is weighted based on its "extremeness", meaning, the more an observation deviates from the linearity, the more it is penalized by giving less weight. The OLS is also just a robust linear regression where all the weights are equal to 1.

There are many common methods to assign the weights to data. Here, we use the package `MASS` from R and its function `rlm`, in which by default the method used for the weights is called "Huber".

The models for each countries are therefore created by using the robust linear regression, and same substitution happens for every model created using linear regression by the original authors (that is, the conditional analysis later on, and plotting the final results and calculation of the resulting coefficients).

Critic on the Gender Equality Index and introduction of the Gender Development Index

The second part of the extended analysis focuses on the index chosen by the authors of the original article to measure the gender (in)equality of the countries. For this, they created a "ad-hoc" gender index (the Gender Equality Index) as a composite measure from four indexes (the WEF Global Gender Gap, the UNDP Gender Inequality Index, the ratio female to male in labor force participation, and the years since women suffrage).

Our analysis started from questioning the choice of these particular indexes (and mostly, the choice of joining them in a single index), based on the doubts raised in (Klasen 2017) about several issues. Here below the most important:

- The Gender Inequality Index from the UNDP is very highly related to the economic development, as its primary weight in the rank is due to the adolescent fertility rate and the maternal mortality;

- Data imputation, as described in the Supplementary Material of this article, but also on the other indexes that impute some values (for instance, the maternal mortality rates) for those countries where such a database is not available or not accurate;

- UNDP Gender Inequality Index has a "inexistent" measure of welfare loss of inequality, because it is based on a calculated gender equality measure that is reported nowhere;

- The WEF Global Gender Gap has a total of fourteen sub-indeces (some of them being the same as for the UNDP GII) where a cap to 1 is applied for those countries where the ratio is higher than 1. This means that countries treating equally men and women are ranked the same as countries disfavoring men.

Following the suggestion in the paper (Klasen 2017), we decided to address some of the problem the author underlined (described above) using the new Gender Development Index from the UNDP report, which is

simply the ratio of the Human Development Index calculated for female divided by the one calculated for male. We added the analysis of this index in the same way as we analysed the other singular indexes, but not adding it to the PCA.

3. Comparison to the Original Article

In this section, we describe how to reproduce the plots and compare the results in terms of z-scores.

Reproducing the Plots of the Main Article

To reproduce the plot of Fig. 1A, we grouped the countries in quartiles based on the logarithm of their average GDP p/c, extracted the mean of each preference from the gender coefficients (the β_1^c) of the countries for each quartile, after standardizing them. The same method was applied to the Gender Equality Index in correlation to the gender differences for each economic preference, to reproduce the plot in Fig. 1C, and to the Gender Development Index, to create the same histogram with a different gender developed countries grouping.

Then, we related the magnitude of the summarised gender difference coefficients (the first component of the PCA) with the logarithm of the average GDP per capita to see the effect of the economic development. This reproduced Fig. 1B of the original article. We used a linear model to fit the correlation and extract the p-value, and for the plot the variables on the y-axis were additionally transformed as (y-y_min)/(y_max-y_min). We then did the same but using a robust linear regression to fit correlation and extracting the p-value.

We applied the same method to extract the correlation between the Gender Equality Index and the summarised gender preference, to see the effect of the gender equality in the countries (Fig. 1D). Note that here also the Gender Equality Index is transformed to be on a scale between 0 and 1. Again, also for this figure we extended the analysis using a robust linear regression to be compared to the original results.

We then extracted the correlation and the p-value of the Gender Development Index to the summarised gender differences, using the robust linear regression (Fig. ExtraGDI). No scale was applied here on the x-axis.

We finally reproduced the plots in Fig. 2A-F using the variable conditioning analysis. This has been done for the economic development, for the Gender Equality Index, and for each of the four indicators building the Gender Equality Index. The variable used on the y-axis is the first Principal Component of the PCA made on the gender differences on the six preferences. All the variables used have been standardize to have mean at 0 and standard deviation of 1 before applying the conditional analysis. Using the residuals, built as described in the Data Analysis section of the Method paragraph, we performed a linear regression on the data points, and we extracted correlation coefficients and p-values.

In the extended analysis, we applied the robust linear regression to the residualisation too. Moreover, we performed the conditional analysis on the Gender Development Index here introduced (Fig. ExtraResidualised)

Tables and z-scores

The comparison of the replication analysis and of the extended analysis to the original article has been done checking the z-scores (calculated using this website: https://www.psychometrica.de/correlation.html) of the correlation coefficients, and comparing the statistical significance. Unfortunately, in the case of the slope coefficients (Table 4) it is not possible to compare them without having the standard deviation of the regression coefficient from the original author study.

Here below we report the tables with the corresponding values of the correlation for the original article, our replication study, the z-scores calculated from them, the correlation extracted using the robust linear regression (extended analysis), and lastly the z-score between the original article and the extended analysis.

In this article, the sample size of the data, needed for the calculation of the z-scores, is 76 (the number of the countries involved in this studies) for the GDP correlations, and 69 for the Gender Equality Index, due to

the missing data. For the original article, the sample size is 76 for the GDP, and 71 for the Gender Equality Index (see Falk & Hermle (2018b), pp. 32, Table S4).

We also indicate the significance level for each correlation using the following scheme:

Significance $\leq$ 0.001 (***), $\leq$ 0.01 (**), $\leq$ 0.05 (*)

Table 1: Correlations between Log GDP p/c and country-level gender differences

Variable	Original	Replication	z-score	Extended	z-score
Altruism	0.58***	0.64***	-0.58	0.62***	-0.38
Trust	0.59***	0.55***	0.36	0.56***	0.27
Positive Reciprocity	0.31***	0.31*	0	0.30*	0.07
Negative Reciprocity	0.35***	0.46***	-0.80	0.49***	-1.03
Risk Taking	0.37***	0.42***	-0.36	0.42***	-0.36
Patience	0.38***	0.43***	-0.36	0.44***	-0.44

Table 2: Correlations between Gender Equality Index and country-level gender differences

Variable	Original	Replication	z-score	Extended	z-score
Altruism	0.51***	0.51***	0	0.45***	0.45
Trust	0.41***	0.48***	-0.51	0.49***	-0.58
Positive Reciprocity	0.13	0.22	-0.54	0.24	-0.66
Negative Reciprocity	0.40***	0.35**	0.34	0.38***	0.14
Risk Taking	0.34***	0.31**	0.19	0.31***	0.19
Patience	0.43***	0.44***	-0.07	0.45***	-0.14

Table 3: Correlation between Log GDP p/c, Gender Equality Index, Gender Development Index, and summarised gender differences

Variable	Original	Replication	z-score	Extended	z-score
Log GDP p/c	0.6685***	0.7119***	-0.484	0.7032***	-0.382
Gender Equality Index	0.5580***	0.5852***	-0.234	0.5754***	-0.148
Gender Development Index	–	–	–	0.3718**	–

Table 4: Conditional analysis to separate the impacts of economic development and gender equality on gender differences

Variable	Residualized on	Slope original	Slope replication	Slope extended
Log GDP p/c	Gender Equality Index	0.5258***	0.5628***	0.5544***
Gender Equality Index	Log GDP p/c	0.3192***	0.2991*	0.2870*
WEF Global Gender Gap	Log GDP p/c	0.2327***	0.2634*	0.2438*
UN Gender Equality Index	Log GDP p/c	0.2911	0.1684	0.1714
F/M in Labor Force Participation	Log GDP p/c	0.2453*	0.2123	0.1949
Years since Women Suffrage	Log GDP p/c	0.2988**	0.1901	0.1990
Log GDP p/c	Gender Development Index	–	–	0.6439***
Gender Development Index	Log GDP p/c	–	–	0.0027

Table 5 (Extra): Comparison of the correlations between Gender Equality Index vs Gender Development Index, and country-level gender differences

Variable	Gender Equality Index (replication)	Gender Development Index (extended)
Altruism	0.51***	0.44***
Trust	0.48***	0.26**
Positive Reciprocity	0.22	0.37***
Negative Reciprocity	0.35**	0.16
Risk Taking	0.31**	0.13
Patience	0.44***	0.19

4. Discussion of the Results

Comparing the original article to the replication analysis

Comparing the results of our analysis to the one from the original paper, starting with the single preferences correlations to the economic development, we see that our analysis brings us to very similar results in terms of correlation coefficients (see Table 1). The p-values, although different, are all indicating a statistically significant correlation, as in the original paper, and when calculating the z-scores thanks to Fisher's r to z transformation, we see that each one is below 2 (which is usually taken as threshold to be statistically significant). This means that our correlations were not statistically significantly different from the ones in the original article.

We do the same for the Gender Equality Index (Table 2), and again we don't find any large difference in the correlation. The p-values tend to be different with respect to the original article, but all of them are pointing towards the same direction. Also the z-scores are in absolute values below 2.

In Table 3, we compare the two core concepts of the article, where the summarised gender differences are regressed on the log GDP p/c and on the Gender Equality Index. The correlations found are similarly positive, strong, and statistically significant. Again, the correlations found in our analysis are not statistically significantly different from the correlations found in the original article.

Lastly, we have the conditional analysis (Table 4). For the two main country-level variable, we see that the values tend to agree and be on the same direction (similar slope coefficients and significant p-value). But when we start to check for the single indexes, we see that there are some differences which are worthy to discuss.

The first thing to say is that we had to make choices on how to impute data and also how to handle the missing data (see discussion above in paragraph "Methods"). The main imputation on missing data has been done on the "time since women's suffrage" dataset, that is where we see a substantial difference in the results. Other datasets, on the other hand, has not been treated for missing data but still they present some difference. For instance, the dataset "F/M in Labor Force Participation" in our analysis has a non-statistically significant correlation, while in the original paper they found a correlation with p-value less than 0.05.

A first thought was that this might be the result of using a different dataset for the GDP (the 2010 USD instead of 2005), but in our opinion this can't be an explanation but rather a check about how robust the results are. So this question about the differences that were found is kept open.

Comparing the original article to the extended analysis

The most interesting part of the analysis arises from the use of the Gender Development Index in place of the Gender Equality index built by the authors. Not only we see very different results in terms of linear dependency for some of the preference (see Table 5), but when we perform the variable conditioning analysis

regressing on the Log GDP p/c, the correlation between gender differences and GDI vanishes (correlation = 0.003, p-value = 0.982).

This brings to the following inputs:

- The measure of the gender differences on the Gender Equality Index is not stable. It strongly depends on the way the gender index is built, and therefore it is hard to claim which one would be a good one for the measure we want to perform. A help can come from the already existing literature (see again [SK]);

- The most important figure to look at is the variable conditioning for the whole analysis, because the gender index is very often intertwined to it, and because in general one can immediately get an answer to the question "how much variance is explained by the variable chosen, independently by the country economic development";

If we look at the results of the two last entries in Table 4, we can deduce that:

1. For countries having the same gender development, the gender differences are depending from the economic development of the country, meaning that richer countries have more gender differences than poorer countries, gender development being equal;

2. 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.

5. 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 replication leads to the same conclusions, but we have some open questions regarding unexplained differences that lead to further checks on the results' robustness.

We decided to focus on two main checks for the robustness: The first was to change the model to have a more "relaxed" assumption for the linear regression, and the choice went to robust linear regression. The results are similar to the ones from the replication analysis, meaning that the model is robust enough.

The second check has been the introduction of a different gender equality indicator, the Gender Development Index (from the UNDP). We performed the same analysis done so far, and we could see some differences rising from it:

- The correlations of the GDI and the gender differences at the single preference level differ quite a lot from the ones obtained using the Gender Equality Index (see Table 5). Half of the gender difference preferences are not correlated significantly anymore to the gender indicator, in one case the correlation is a half of the original one, in another case it starts to be significant (while in the original it wasn't). Only for the "altruism" we see the same behavior;

- The correlation between GDI and gender differences is positive and statistically significant, although the strength of the correlation is somewhat less than the one for the GEI used by the authors;

- Lastly, the most interesting result is that the gender differences residualised on the economic develooment don't show a dependency on the GDI (also residualised on economic development), while the vice-versa applies (meaning, there is a strong and statistically significant correlation between gender differences and economic development, both residualised on the GDI). This result arises many questions that are beyond the scope of this analysis.

References

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.

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.

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.

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

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.