

Appendix – Replicate and extend the results of the article “Relationship of gender differences in preferences to economic development and gender equality”

Sara Cerioli

Andrey Formozov

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.

The following packages with respective version are used:

Package	Version
data.table	1.13.2
bit64	4.0.5
bit	4.0.4
plyr	1.8.6
dplyr	1.0.5
haven	2.4.2
ggplot2	3.3.2
missMDA	1.18

Data Collection, Cleaning, and Standardization

The data used by the authors is not fully available because of two reasons:

1. **Data paywall:** Some sociodemographic variables (for instance, education level or income quintile) are not part of the Global Preference Survey, but of the Gallup World Poll dataset. Check the website of the briq - Institute on Behavior & Inequality for more information on it.
2. **Data used in study is not available online:** This is what happened for the log GDP p/c calculated in 2005 US dollars (which is not directly available online). We decided to calculate the log GDP p/c in 2010 US dollars because it was easily available, which should not change the main findings of the article.

An additional issue that we faced while trying to reproduce the results of the article has been the missing data. We will treat this specific issue later on because it requires a bit of background.

The procedure for cleaning is described for each dataset in the corresponding section below. After manually cleaning the dataset, we standardized the names of the countries and merged the datasets into one.

Global Preferences Survey

This data is protected by copyright and can't be given to third parties.

To download the GPS dataset, go to the website of the Global Preferences Survey in the section "downloads". There, choose the "Dataset" form and after filling it, we can download the dataset.

GDP per capita

From the website of the World Bank, one can access the data about the GDP per capita on a certain set of years. We took the GDP per capita (constant 2010 US\$), made an average of the data from 2003 until 2012 for all the available countries, and matched the names of the countries with the ones from the GPS dataset.

Gender Equality Index

The Gender Equality Index is composed of four main datasets. Here below we describe where to get them (as originally sourced by the authors) and how we treated the data within them, if needed.

- **Time since women's suffrage:** Taken from the Inter-Parliamentary Union Website. We prepared the data in the following way. For several countries more than one date were provided (for example, the right to be elected and the right to vote). We use the last date when both vote and stand for election right were granted, with no other restrictions commented. Some countries were colonies or within union of the countries (for instance, Kazakhstan in Soviet Union). For these countries, the rights to vote and be elected might be technically granted two times within union and as independent state. In this case we kept the first date. It was difficult to decide on South Africa because its history shows the racism part very entangled with women's rights [citation]. We kept the latest date when also Black women could vote. For Nigeria, considered the distinctions between North and South, we decided to keep only the North data because, again, it was showing the completeness of the country and it was the last date. Note: USA data doesn't take into account that also up to 1964 black women couldn't vote (in general, Blacks couldn't vote up to that year). We didn't keep this date, because it was not explicitly mentioned in the original dataset. This can be seen as in contrast with other choices made though.
- **UN Gender Inequality Index:** Taken from the Human Development Report 2015. We kept only the table called "Gender Inequality Index".
- **WEF Global Gender Gap:** WEF Global Gender Gap Index Taken from the World Economic Forum Global Gender Gap Report 2015. For countries where data was missing, data was added from the World Economic Forum Global Gender Gap Report 2006. NOTE: We modified some of the country names directly on the csv file, that is why we provide this as an input file.
- **Ratio of female and male labour force participation:** Average International Labour Organization estimates from 2003 to 2012 taken from the World Bank database (<http://data.worldbank.org/indicator/SL.TLF.CACT.FM.ZS>). Values were inverted to create an index of equality. We took the average for the period between 2004 and 2013.

Missing Data and Imputation

During the reproduction of the article, we found that the authors didn't write in details how they handled missing data in the indicators.

They mention on page 14 of the Supplementary Material, that (quoting): "For countries where data were missing data were added from the World Economic Forum Global Gender Gap Report 2006 (http://www3.weforum.org/docs/WEF_GenderGap_Report_2006.pdf)."

However, there are two problems here:

- Regarding the year when women received the right to vote in a specific country. The missing values are the ones coming from the United Arab Emirates and Saudi Arabia, that neither in 2006 (when the

WEF Global Gender Gap Report that the authors quote as a reference for the missing values) nor now (in 2021) have guaranteed the right to vote for women yet.

- There are missing data also in the other sources that the authors quote. So a quick search for the missing countries of the WEF report of 2015, shows us that these countries can't be found in the report of 2006 either.

These two unclear points, even though in our understanding not crucial for the replication of the analysis, are not desirable, but couldn't be further clarified with the authors.

The problem of missing data for a given countries often does not influence much the overall trends of found correlations. However, it is very relevant if one would to see the implications of the study with respect to a specific country of interest.

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. We used in addition a Robust Linear Regression after checking the normality of the data.
- 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.

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 2018) article addresses the gender differences, the main focus is on that individual variable and all the others provided in the dataset (education level, income quintile, 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 kind of plots helps determining the normality of the residuals by looking at their distribution along a straight diagonal line.

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).

Principal Component Analysis

To summarise the average gender difference among the six economic preferences, we performed a principal component analysis (PCA) 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 maximise the variance. The first component of the PCA has then been used as a summary index of average gender differences in preferences.

We performed a PCA also on the four datasets used for Gender Equality, to extract a summarised Gender Equality Index.

Variable Conditioning

To separate the effects of the economic development and the gender equality, a conditional analysis was performed (Frisch & Waugh (1933), and Lovell (1963)). To generalise, if one wants to estimate the correlation of x and y conditioning on z , one needs to perform a double linear regression:

- First, regressing x on z and extracting the residuals
- Second, regressing y on z , and extracting the residuals.

In the end, one needs to take the so calculated residuals of x on z and of y on z , and make a last regression to calculate the correlation.

In practice, if we are interested in checking the influence of the economic development on the summarised gender differences, conditioning on the gender equality, we would need to regress the economic development on the gender equality index, then the average gender differences regressed on the gender equality index, and finally regress the residuals of the average gender differences on the residuals of the economic development.

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.

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 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.

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.

Tables and z-scores

Comparing the results of our analysis to the one from the original paper, starting with the summarised gender preferences to the economic development and the gender equality, we see that our analysis brings us to very similar results in terms of correlation coefficients (see Table 1). The p-values 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. Very similar results appear also from the robust linear regression used in the extended analysis. We do the same for the Gender Development Index, and also here we find a positive, statistically significant correlation, even if slightly less large than the one found for the Gender Equality Index.

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

Significance ≤ 0.001 (***), ≤ 0.01 (**), ≤ 0.05 (*)

Table 1: Correlation between PCA-summarised gender differences in economic preferences vs log GDP p/c, aggregated Gender Equality Index, and Gender Development Index. The agreement between this study and the original is quantified as *z-score*.

	Log GDP p/c	Gender Equality Index	Gender Development Index
Original	0.6685***	0.5580***	–
Replication	0.7119***	0.5852***	0.3929***
<i>z-score</i>	-0.484	-0.234	–
Extended	0.7032***	0.5754***	0.3718**
<i>z-score</i>	-0.382	-0.148	–

In Table 2, we summarise the results of the correlation of single preferences to the gender equality of the countries. We can see very different results in terms of linear dependency for some of the preferences when comparing the correlation coming from the Gender Equality Index and the one coming from the Gender Development Index. Note that for the Gender Equality Index, we used the results obtain by this replication study, in order to have a more consistent comparison (additionally, we have already seen that there are no substantial differences from this study and the original article, see again Table 1).

Table 2: Comparison of the correlations between Gender Equality Index vs Gender Development Index, and country-level gender differences in economic preferences. The correlations are calculated using the OLS method.

Variable	Gender Equality Index	Gender Development Index
Altruism	0.51***	0.47***
Trust	0.48***	0.25*
Positive Reciprocity	0.22	0.37**
Negative Reciprocity	0.35**	0.14
Risk Taking	0.31**	0.15
Patience	0.44***	0.21

Lastly, in Table 3, we summarise the results of the conditional analysis. For the two main country-level variables, 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.

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. When the variable conditioning analysis regressing on the Log GDP p/c is done, the correlation between gender differences and GDI vanishes (correlation = 0.0027, p-value = 0.982).

Add the plots of extended analysis with Log GDP residualize by GDI and viceversa

Table 3: Conditional analysis to separate the impacts of economic development and gender equality on gender differences in economic preferences. Reported are the slopes of the linear regressions.

Variable	Residualized on	Original	Replication	Extended
Log GDP p/c	Gender Equality Index	0.5258***	0.5673***	0.5657***
Gender Equality Index	Log GDP p/c	0.3192***	0.2856*	0.2972*
WEF Global Gender Gap	Log GDP p/c	0.2327***	0.2006*	0.1917*
UN Gender Equality Index	Log GDP p/c	0.2911	0.2355 ^a	0.2385
F/M in Labor Force Participation	Log GDP p/c	0.2453*	0.1708 ^b	0.1684
Years since Women Suffrage	Log GDP p/c	0.2988**	0.1499 ^c	0.1561
Log GDP p/c	Gender Development Index	–	0.7055***	0.7214***
Gender Development Index	Log GDP p/c	–	0.0307	0.0020

^aLikely due to update on the webpage

^bNot understood

^cArbitrary retrieval of the data (check also Appendix)

Table S1: Correlation coefficients for country-level gender differences in economic preferences vs log GDP p/c obtained in the original article, present replication and extended analyses. The agreement between results is quantified as *z-score*. [Check the merging procedure, there are missing countries]

Economic preference	Original analysis	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	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 S2: Correlation coefficients for country-level gender differences in economic preferences vs aggregated Gender Equality Index obtained in the original article, present replication and

extendent analyses. The agreement between results is quantified as *z-score*.

Economic preference	Original	Replication	z-score	Extended	z-score
Altruism	0.51***	0.51***	0.0	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

References

- Falk, A. & Hermle, J. (2018). Relationship of gender differences in preferences to economic development and gender equality. *Science*, **362**.
- Frisch, R. & Waugh, F.V. (1933).
- Lovell, M.C. (1963). Seasonal adjustment of economic time series and multiple regression analysis. *Journal of the American Statistical Association*, **58**, 993–1010.