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. 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

Global Preferences Survey and Gallup World Poll datasets

To download the GPS dataset, one can go to the website of the Global Preferences Survey briq - Institute on Behavior & Inequality in the section "downloads". There, choose the "Dataset" form and after filling it, one can download the dataset. 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 that is not openly availible. This data is protected by copyright and can't be given to third parties. Check the website of the briq - Institute on Behavior & Inequality for more information on it.

GDP and gender equality indexes

From the website of the World Bank, one can access the data about the GDP per capita on a certain set of years. The data for log GDP p/c calculated in 2005 US dollars was already archived. We used log GDP p/c in 2010 US dollars, instead. To build an estimator for Log GDP p/c, we averaged 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.

The Gender Equality Index used in the original article was composed of four main datasets as the first principle component of the PCA of them:

- **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.

- **UN Gender Inequality Index:** Taken from the Human Development Report 2015. We kept only the table called "Gender Inequality Index".

- **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/indicat or/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.

- **Time since women's suffrage:** Taken from the Inter-Parliamentary Union Website. For several countries more than one date where 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.

In this work we additionally involve the data GDI index:

- [**Add link to GDI here**]

Missing Data and Imputation

The procedure for imputation and "cleaning"" for each dataset is described in the corresponding section below. e standardized the names of the countries and merged the datasets into one.

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.

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 [check the right to vote, actually there is some progress], 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 [are there this points on the final plots?]. 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. Missing data and imputation in general are not crucial for the replication of the analysis, yet not desirable. 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 like 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 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 used the diagnostic plots for the linear regression to verify the normality of the data, 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 a joint 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); 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.

Pure replication and 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 replication (using linear regression) and extended (with robust linear regression) analysis to the one from the original paper (see Falk & Hermle (2018a) Fig. 1B and 1D), 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.

Table 2: Correlation between PCA-summarised gender differences in economic preferences vs log GDP p/c, aggregated Gender Equality Index, and Gender Development Index. The number of countries in each dataset was 76. The agreement between results is quantified as *z-score*. Significance ≤ 0.001 (***), ≤ 0.01 (**), ≤ 0.05 (*)

		Log GDP p/c	Gender Equality Index
Original		0.6685***	0.5580***
Replication		0.6830***	0.6079***
	z-score	-0.161	-0.457
Extended		0.6733***	0.5905***
	z-score	-0.053	-0.293

The comparison of the Fig. 2A-F of Falk & Hermle (2018a) is done in Table 1 of the main article, where we compare the conditional analysis results for original, replicated and extended study. Reported are the slopes of the linear regressions and the corresponding p-value.

In Table 2, we summarise the results of the correlation of single preferences to the economic development of the countries, to be compared to Fig. S2 in Falk & Hermle (2018b), while in Table 3, we show the results for the single preferences gender differences regressed on the Gender Equality Index, as done in Fig. S3 in Falk & Hermle (2018b).

Table 3: Correlation coefficients for country-level gender differences in economic preferences vs log GDP p/c obtained in the original article, present replication and extended analysis. The agreement between original and this work is quantified as *z-score*. Significance ≤ 0.001 (***), ≤ 0.01 (**), ≤ 0.05 (*)

		Trust	Altruism	Pos. Rec.	Neg. Rec.	Risk Tak.	Patience
Original		0.5918***	0.5847***	0.3086**	0.3542**	0.3685**	0.3837***
Replication		0.5761***	0.5882***	0.3216**	0.3683**	0.3649**	0.4415***
	z-score	0.144	-0.032	-0.087	-0.098	0.025	-0.421
Extended		0.5926***	0.5488***	0.3240**	0.3929***	0.3579**	0.4253***
	z-score	-0.007	0.32	-0.103	-0.272	0.074	-0.301

Table 4: 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. The agreement between original and this work is quantified as *z-score*. Significance ≤ 0.001 (***), ≤ 0.01 (**), ≤ 0.05 (*)

		Trust	Altruism	Pos. Rec.	Neg. Rec.	Risk Tak.	Patience
Original		0.4050***	0.5073***	0.1280	0.4035***	0.3412**	0.4257***
Replication		0.4829***	0.5030***	0.2132	0.3751***	0.3441**	0.4415***
	z-score						
Extended		0.4921***	0.4686***	0.2156*	0.3714**	0.3414***	0.4434***
	z-score						

For the comparison of the results from Fig. S4 in Falk & Hermle (2018b), we refer to Table 4, showing the correlation between the average gender differences to the single gender equality indicators.

Table 5: Single indicator for the gender equality correlated with gender differences. Significance ≤ 0.001 (***), ≤ 0.01 (**), ≤ 0.05 (*)

		WEF GGGI	UNDP GII	F/M LFP	Time Since Women Suffrage
Original		0.4097***	0.6482***	0.2661*	0.5139***
Replication		0.4075***	0.6680***	0.2860*	0.4517***
	z-score				
Extended		0.3937**	0.6586***	0.2645*	0.4535***
	z-score				

Comparing the Fig. S5 and S6 of Falk & Hermle (2018b) to our results: See Table 5 and Table 6.

Table 6: Log GDP p/c conditional on Gender Equality Index and on Gender Development Index, regressed on each single economic preference gender difference coefficient. Here reported are the correlation terms for the original analysis, the replication (using OLS) and the extended analysis (robust linear regression). Significance levels: ≤ 0.001 (***), ≤ 0.01 (**), ≤ 0.05 (*).

		Trust	Altruism	Pos. Rec.	Neg. Rec.	Risk Tak.	Patience
Original		0.4574***	0.4751***	0.2771*	0.2444*	0.2868*	0.2621*
Replication		0.4265***	0.4338***	0.2509	0.2111	0.2256*	0.2288*
	z-score						
Extended		0.4450***	0.3959***	0.2524	0.2451*	0.2191*	0.2444*
	z-score						

Table 7: Gender Equality Index and Gender Development Index, conditional on Log GDP p/c, regressed on each single economic preference gender difference coefficient. Here reported are the correlation terms for the original analysis, the replication (using OLS) and the extended analysis (robust linear regression). Significance levels: ≤ 0.001 (***), ≤ 0.01 (**), ≤ 0.05 (*).

		Trust	Altruism	Pos. Rec.	Neg. Rec.	Risk Tak.	Patience
Original		0.2050	0.3304**	-0.0115	0.2788*	0.1973	0.2967*
Replication		0.2472*	0.2696*	0.0481	0.2240*	0.1863*	0.2841*
	z-score						
Extended		0.2521*	0.2401	0.051	0.2037*	0.1868*	0.2781*
	z-score						

Table 8: Preferences standardized at global level for Log GDP p/c.

		Trust	Altruism	Pos. Rec.	Neg. Rec.	Risk Tak.	Patience
Original		0.5787***	0.3304**	0.2819*	0.2980**	0.2974**	0.4391***
Replication		0.5761***	0.5882***	0.3216**	0.3683**	0.3649**	0.4124***

	Trust	Altruism	Pos. Rec.	Neg. Rec.	Risk Tak.	Patience
z-score						

Table 9: Gender differences and economic development by preference and country without controls.

	Trust	Altruism	Pos. Rec.	Neg. Rec.	Risk Tak.	Patience
Original	0.5434***	0.5808***	0.2748*	0.4038***	0.3860***	0.4830***
Replication	0.5462***	0.5881***	0.2826*	0.3864***	0.3886***	0.4804***
z-score						

For the comparison with Fig. 1A, 1B, S1, we have reproduced the plots but a comparison of the values wasn't possible, as no value is provided. So one can compare "by eyes".

For the replication of the Fig. S7, we couldn't replicate because we had no access to such granularity of the data.

For the replication of the tables in the supplementary material, we were not able to get the dataset as indicated from the sources.

References

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

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