

# Facebook's gender divide

David Garcia,<sup>a,b</sup> Yonas Mitike Kassa,<sup>c,d</sup> Angel Cuevas,<sup>d</sup>  
 Manuel Cebrian,<sup>e,f</sup> Esteban Moro,<sup>g</sup> Iyad Rahwan,<sup>f,h</sup> Ruben Cuevas<sup>d</sup>

<sup>a</sup> Complexity Science Hub Vienna, Vienna, Austria

<sup>b</sup> Medical University of Vienna, Vienna, Austria

<sup>c</sup> IMDEA Networks Institute, Madrid, Spain

<sup>d</sup> Universidad Carlos III de Madrid, 28911 Leganes, Spain

<sup>e</sup> Data61, CSIRO, Melbourne, Australia

<sup>f</sup> The Media Lab, Massachusetts Institute of Technology, Cambridge MA, USA

<sup>g</sup> Department of Mathematics & GISC,

Universidad Carlos III de Madrid, 28911 Leganes, Spain

<sup>h</sup> Institute for Data, Systems & Society,

Massachusetts Institute of Technology, Cambridge MA, USA

October 11, 2017

## Abstract

Online social media are information resources that can have a transformative power in society. While the Web was envisioned as an equalizing force that allows everyone to access information, the digital divide prevents large amounts of people from being present online. Online social media in particular are prone to gender inequality, an important issue given the link between social media use and employment. Understanding gender inequality in social media is a challenging task due to the necessity of data sources that can provide unbiased measurements across multiple countries. Here we show how the Facebook Gender Divide (FGD), a metric based on a dataset including more than 1.4 Billion users in 217 countries, explains various aspects of worldwide gender inequality. Our analysis shows that the FGD encodes gender equality indices in education, health, and economic opportunity. We find network effects that suggest that using social media has an added value for women. Furthermore, we find that low values of the FGD precede the approach of countries towards economic gender equality. Our results suggest that online social networks, while suffering evident gender imbalance, may lower the barriers that women have to access informational resources and help to narrow the economic gender gap.

The Web was designed to be universally accessible and open, carrying the promise of equal opportunity in the access to online information and services [4] as *the great potential equalizer* [14]. However, despite the widespread adoption of the Web and other Information Communication Technologies (ICT), online access and presence are heterogeneous across demographic factors such as income and gender –a phenomenon called the *digital divide* [5, 8, 19].

Governments and global organizations express their concern about the digital divide, aiming to connect the 4 Billion people that remain offline [23, 13]. However, the effects of increasing Internet penetration in development are rarely backed up against empirical data [10], and the latest report by the World Bank suggests that unequally distributed growth in Internet penetration might exacerbate socio-economic inequalities [13].

Among online resources, social media play a key role in economic development, for example by providing information that facilitates finding employment [6, 12]. An important open question is whether equality in the access to social media can work as a digital provide [16], bringing equality in other social, political, and economic aspects of society.

The World Wide Web Foundation reports that one of the key elements in the digital divide is gender inequality [26]. Gender-based online violence and sexual harassment are pervasive phenomena across countries and can drive gender inequality in activity even when access is equally available across genders, generating gender divides that demand research [26]. Empirical analyses of social media data show the traces of gender inequalities, from content biases and activity on Wikipedia [25, 21] to visibility and interaction disparities on Twitter [11, 18].

To properly understand the digital divide, a pervasive problem in cross-country comparisons is the limited size of country samples and the challenges to generate unbiased survey data [13]. To overcome this issue, we deployed a novel technical infrastructure to collect large-scale data from the Facebook online social network, as explained more in detail in the Materials and Methods section. We generated a dataset with the total amount of registered users and daily active users of each gender in each country, covering a total of 217 countries and more than 1.4 Billion users (more details in Supplementary Information).

Our dataset allows the quantification of Facebook activity ratios of each gender in each country. From them, we calculate the *Facebook Gender Divide* (FGD) as the logarithm of the ratio between the activity ratios for men and for women (see Methods for more details). The FGD has a value below zero when women tend to be more active on Facebook than men, a value close to zero for equal activity tendencies, and a positive value when men are more active on Facebook than women in a country. We combine this data with standard socio-economic indices, including Gross Domestic Product (GDP) per capita, Internet penetration (IP), and economic inequality, as well indices from the World Economic Forum Gender Gap Report that measure gender equality in terms of education, health, political participation, and economic opportunities.

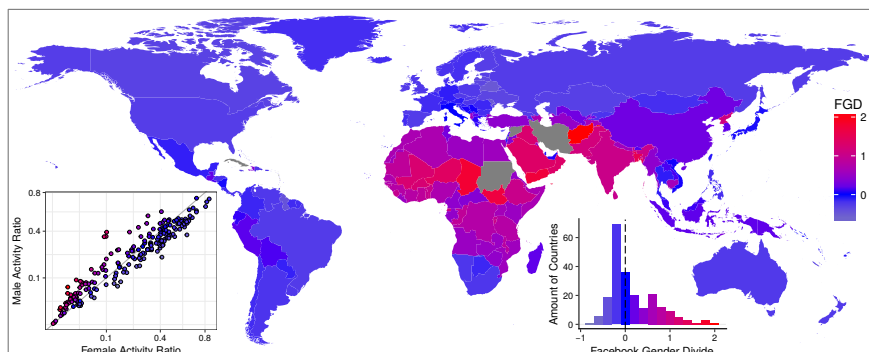


Figure 1: **The Facebook Gender Divide across 217 countries.** Countries are colored according to their Facebook Gender Divide (FGD), from highly skewed towards males (red), balanced (blue), and towards females (green, not visible). The left inset shows the scatter plot of male and female activity ratios across all countries, revealing a spread along the diagonal. The right inset shows the histogram of FGD values in bins of width 0.2. While the mode of countries is slightly below zero, there is significant skewness towards high FGD values.

## Results

Fig. 1 shows a world map with countries colored according to their FGD, revealing that many countries are very close to gender equality in Facebook (blue color). The red scale shows countries with positive FGD—that is, higher proportion of males on Facebook. The scale towards FGD below zero (more tendency for women to be on Facebook) is much narrower than above zero, as can be seen in the scatter plot with the activity ratios of each gender (Fig. 1, left inset), and in the skewness of the distribution of FGD across countries (Fig. 1, right inset). An online interactive version of Fig. 1 can be found in <http://dgarcia.eu/FacebookGenderDivide>.

Countries with strong FGD above zero are located around Africa and South West Asia, as shown on Fig. 1. This suggests that variations in socio-economic factors of gender inequality across regions could be explanatory of the FGD. We test this observation using a linear regression model of the FGD as a function of the four indices of gender equality measured by the World Economic Forum (economic opportunity, education, health, and political participation), plus three non-gender-based controls of Internet penetration, population size, and economic inequality (see Methods). The left panel of Fig. 2 shows the quality of the model fit, comparing empirical values of FGD rank versus model predictions. Remarkably, the model can explain well the ranking of FGD ( $R^2 = 0.74$ ), with very few points far from the diagonal.

The right panel of Fig. 2 shows the estimate of the coefficients of our model of FGD. The strongest coefficient is that of education gender equality, which can also be observed on the colors of the left panel of Fig. 2. Specifically,

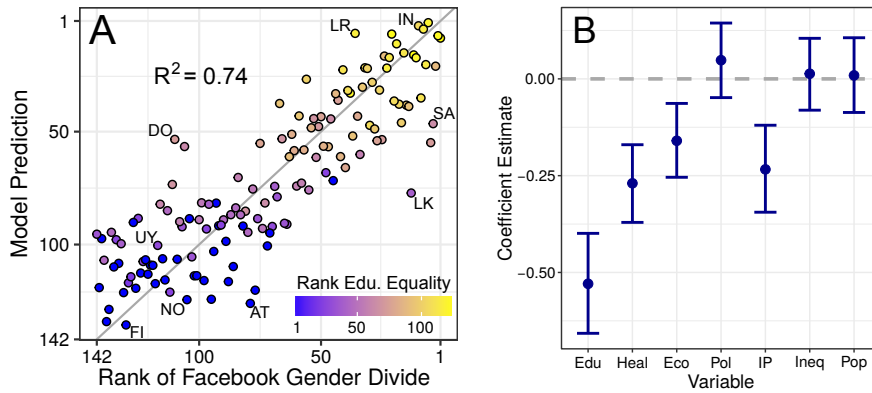


Figure 2: **Regression results of FGD as a function of gender equality.** A) Model predictions versus rank of FGD, where rank 1 is the country with the highest FGD. The model achieves a high  $R^2$  above 0.74, explaining the majority of the variance of the FGD ranking. Some countries are labelled, from high FGD (Liberia, India, and Saudi Arabia) to low FGD (Finland, Norway, and Uruguay), as well as some outliers (Dominican Republic, Austria, and Sri Lanka). B) Coefficient estimates and 95% CI of the terms of the regression fit (excluding intercept). Education (Edu), health (Heal), and economic gender equality (Eco) are significantly and negatively associated with the FGD, but political gender equality (Pol) is not. From the control variables, Internet penetration (IP) is negatively associated with FGD, but economic inequality (Ineq) and population (Pop) are not. The main role of education equality in FGD can be observed on panel A, where dots are colored according to the rank of education gender equality, showing that countries with low FGD are ranked high on education gender equality. An online interactive version of this figure can be found in <http://dgarcia.eu/FacebookGenderDivide>

countries with high rank in this index have, on average, lower FGD. Health and economic gender equality also have significant negative coefficient estimates, showing that the FGD captures more than one type of inequality. Note that the index for political gender equality does not have a significant relationship with FGD when the other indices are considered in the model. Among controls, only Internet penetration is negatively associated with FGD. This also appears in similar regression models including GDP and other development indices, evidencing that the relationship between gender equality indices and FGD is observable when development metrics are considered. We present these additional controls, regression diagnostics, and robustness tests in Supplementary Information, concluding that the negative relationships between FGD and gender equality indices are robust.

The value of being active in social media might vary across genders, which we can analyze in a wide country comparison. Network effects are a possibility in social media, accelerating activity rates as a social networking technology

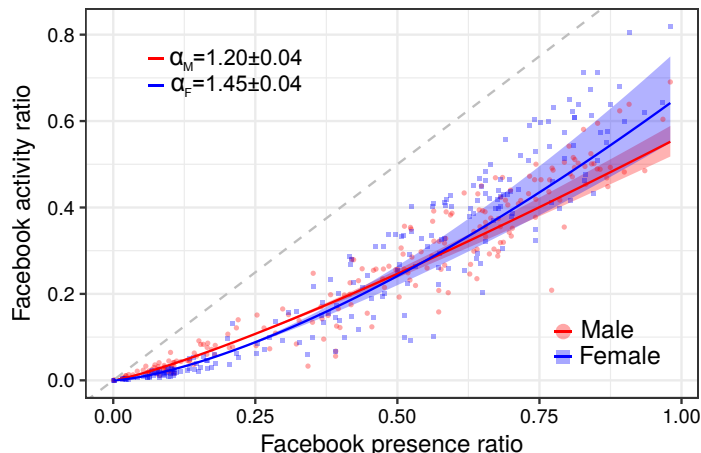


Figure 3: **Gender differences in network effects on Facebook.** Scaling of Facebook activity ratio versus presence ratio. Solid lines show fit results and shaded areas show their 95% confidence intervals. Both male and female activity ratios grow superlinearly with presence ( $\alpha > 1$ ), indicating a network effect. This network effect is stronger among female than among male users.

penetrates a society. An example of this is Metcalfe’s law [15], by which the individual value of using a communication medium grows with the amount of people connected to it. If there is a network effect, the activity ratio in a country should grow superlinearly with the ratio of people with an account in the country, regardless of whether they are active or not. Furthermore, this effect might vary for male and female users, revealing whether using Facebook might have higher marginal benefit for one gender than for the other.

Fig. 3 shows the scaling relationship per gender between the activity ratio and the total Facebook presence ratio in each country. Lines show the result of a power-law fit between both variables with an intercept and an interaction term for gender. The estimate of the scaling exponent for each gender is clearly above one for both genders, revealing a superlinear trend consistent with a network effect in Facebook. This exponent is significantly stronger for female users ( $\alpha_F = 1.45[1.41, 1.49]$ ) than for male users ( $\alpha_M = 1.20[1.16, 1.24]$ , see Supplementary Information for details). This result suggests a surprising phenomenon, namely that the *utility of being active on Facebook might be higher for women*. A possible explanation is that social media allows women in countries with high gender inequality to access social capital that might not be accessible through other means—e.g. due to restricted spatial mobility [24].

We further test the possibility that low levels of FGD are related to narrowing gender gaps in other aspects of society. We fitted a model of changes in economic gender equality as a function of the rank of FGD in the previous year, including controls for autocorrelation and economic factors. The coefficient estimates, shown in Fig. 4 and explained more in detail in Supplementary Information,

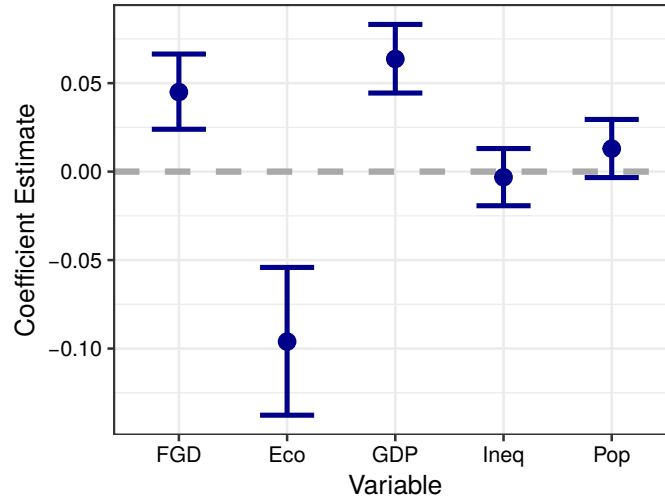


Figure 4: **The role of the FGD on changes in economic gender equality.** Coefficient estimates of the regression model of changes in economic gender equality as a function of FGD and control terms (excluding intercept). A low rank of FGD is associated with changes towards economic gender equality.

reveal a significant positive relationship between the FGD rank and changes in economic gender inequality. This means that countries with low values of FGD (i.e. high rank number) tend more to approach economic gender equality. On the contrary, this association is not observable on education nor political gender inequality (see Supplementary Information), suggesting that social media plays a role in bridging economic gaps but not in other social or political inequalities. The same result holds when we include education equality in the model of changes in economic gender equality. In combination with the results of the FGD model presented in Fig. 2, these results suggest a novel development sequence in gender equality: education gender equality is associated with gender equality in social media activity, which in turn leads to increasing levels of economic gender equality.

## Discussion

By quantifying the Facebook Gender Divide among 1.4 Billion Facebook users, we demonstrate a number of phenomena that deserve further investigation. The FGD is associated with other types of gender inequality, including economic, health, and education inequality. While the mechanisms behind this connection remain an open question, this work is an example of how publicly accessible social media data can be used to understand an important social phenomenon. The FGD provides an inexpensive and accessible way to compute gender di-

vides in social media that can be tracked over time and across countries in development analyses [22]. Future studies can include the FGD along with other indicators to understand the role of ICT in individual countries, taking into account their particularities before formulating policy suggestions.

We found evidence of a gender-dependent network effect—that is, women might receive higher marginal benefit than men from the general adoption of Facebook in a country. Furthermore, we identified a relationship between gender equality in Facebook and changes in economic gender inequality in which countries with lower Facebook Gender Divide approach faster economic gender equality. These results suggest that social media can be an equalizing force that counteracts other barriers—e.g. those that limit women’s mobility—by providing access to greater economic opportunities and social capital. In a similar way as mobile phones increased the life quality of fishermen in India [16], social media might work as a *digital provide* that helps disfavoured groups despite the still generalized inequalities in access to ICT and in adoption of social media technologies.

## Materials and Methods

### The Facebook Global Dataset

We collected the number of Facebook users by age and gender in each country using the Facebook graph API [1]. Among other services, this API delivers marketing data for its commercial customers to provide targeted advertising. When supplied with a specific target population, the API returns the total audience size and the price to reach that target audience through Facebook. We iterated over each combination of age and gender values, retrieving the total amount of users and the amount of Daily Active Users (DAU) for each segment in each country. Our dataset contains the number of male and female registered users and daily active users for all the countries available in the graph API<sup>1</sup>. After removing entries of small countries with missing values, our dataset contains the total amount of users and DAU segmented by age and gender for 217 countries. Age data in the API starts at 13, increasing by one year up to a last bin that contains all users aged 65 or older. To correct any possible fluctuation on the data reported by Facebook, we repeated our analysis on monthly snapshots of Facebook data for a period of twelve months between 2015 and 2016, as reported in Supplementary Information.

### Gender equality and development datasets

To normalize the amount of active users over the total population of each country, we use the data collected by the US Census Bureau. This dataset contains estimates of the resident population by age and gender for more than 226 countries. We combine this data with gender equality indices measured by the World

---

<sup>1</sup>The API does not deliver data for certain countries, e.g. Syria, Iran, and Cuba.

Economic Forum Gender Gap reports of 2015 and 2016 [3]. This dataset quantifies the magnitude of gender equality in 145 countries, measuring it with respect to four key areas: health, education, economic opportunity, and politics. This report updates the values for education, economic, and political gender equality on a yearly basis, allowing us to measure changes between 2015 and 2016. To account for additional economic and development indicators, we include data from the World Bank and the Human Development Index [2], measuring control variables of GDP (PPP) per capita in 2012, economic inequality as the quintile ratio, and Internet penetration.

## Computing the Facebook Gender Divide

We quantify the Facebook Gender Divide as a comparison of the rates of activity between genders. The amount of Daily Active Users (DAU) measures how many users have logged into Facebook at a given day, which could be either through a Web browser or a mobile application. We use the segmented data from 13 to 65 years old to normalize the amount of DAU over the total population of a country in those ages, truncating all data that is not included in that age range. This way we avoid introducing a bias with life expectancy and average age. To have a stable estimate of the DAU, we calculate the median value over the month of July 2015. We repeated our analysis with similar measurements over the following year, validating that our results do not depend on the choice of that particular time period. This way, for each country  $c$  and gender  $g \in [Female, Male]^2$ , we have a measurement of the amount of active users  $A_{g,c}$  between 13 and 65 years old.

Using the US Census Bureau data, we calculate the total population of each gender between the ages of 13 and 65 years old in each country, which we denote as  $P_{g,c}$ . This way, we can normalize the total activity in Facebook over the population in the same age ranges, calculating the ratios  $R_{g,c}$ . We define the **Facebook Gender Divide** in country  $c$  as

$$FGD_c = \log \left( \frac{R_{Male,c}}{R_{Female,c}} \right)$$

, which compares male and female Facebook activity rates over the population of country  $c$ . A country with positive FGD will have a tendency for men to be more present on Facebook, while a country with negative FGD will show the opposite tendency. A country with  $FGD = 0$  will have complete equality in the activity tendencies of both genders.

## Regression Models

We model dependencies between gender equality indicators and the FGD as linear models, after applying a rank transformation to all variables such that rank

---

<sup>2</sup>For simplicity, we take gender as birth sex, i. e. male or female.



1 is the highest possible value of the variable. This way, we explore monotonic dependencies that do not need to be linear. We define this *FGD model* as:

$$FGD = a_f \cdot Q + b_f \cdot C + c_f + \epsilon$$

where  $Q$  is a matrix with the ranks of economic, health, education, and political gender equality in each country and  $C$  contains control variables such as Internet penetration, income inequality, and total population.  $c_f$  is the intercept and  $\epsilon$  denotes the residuals as the normally distributed, uncorrelated error of the model.

We analyze the relationship between changes in economic gender equality and the rank of FGD through the following *equality changes model*:

$$\Delta Eco = a_o \cdot FGD + b_o \cdot Eco_{2015} + c_o \cdot O + d_o + \psi$$

where  $\Delta Eco$  is the change in economic gender inequality between 2015 and 2016. The inclusion of the term  $Eco_{2015}$  controls for autocorrelation in the economic gender equality index. Following previous Economics research on Facebook data [12], we include the control matrix  $O$  with economically relevant rank-transformed variables: GDP, Income inequality, and the total population of each country.  $d_o$  is the intercept of the model and  $\psi$  denotes the residuals.

We model network effects as a power-law relationship between the activity ratio of a gender and the total Facebook presence ratio, where the Facebook presence ratio of country  $c$ , denoted as  $R'_c$ , is the ratio between the total amount of accounts (active or inactive) in country  $c$  over its total population. We define this *network effects model* as:

$$\log(R_{g,c}) = \alpha \cdot \log(R'_c) + \beta + \delta_{g,Female}(\alpha_F \cdot \log(R'_c) + \beta_F) + \phi$$

where  $\alpha$  measures the scaling relationship between the Facebook presence ratio and the activity ratio of male users,  $\alpha_F$  the difference in that relationship for female users, and  $\phi$  the residuals. The Kronecker delta function  $\delta_{g,Female}$  takes value 1 when  $g = Female$  and 0 otherwise.

The above models do not show relevant multicollinearity, as revealed by the calculation of Variance Inflation Factors [7] reported in the Supplementary Material. We fit all models using Markov Chain Monte Carlo sampling in JAGS [20]. We assume a narrow non-informative Normal prior around zero for all parameters, and we calculate 95% Credible Intervals (CI) over the posterior distribution of each parameter to assess the uncertainty of median estimates.

To test the validity of the assumptions of our models after fitting, we verify the normality of residuals through Shapiro-Wilk tests [9], and check that residuals are uncorrelated with fitted values and independent variables. We further compare the fit results against the outcome of robust regression through the Method of Moments [17], validating that our results are robust to the possible existence of outliers. For the case of the network effects model, we additionally analyze multiplicative residuals in order to test for the possible role of outliers, as shown more in detail in Supplementary Information.

## Acknowledgements

D.G. acknowledges funding from the Vienna Science and Technology Fund through the Vienna Research Group Grant “Emotional Well-Being in the Digital Society” (VRG16-005). E.M. acknowledges funding from Ministerio de Economía y Competividad (Spain) through projects FIS2013-47532-C3-3-P and FIS2016-78904-C3-3-P. A.C. acknowledges funding from the H2020 EU project TYPES (grant no. 653449) and the Ramón y Cajal grant (RyC-2015-17732). Y.M.K. acknowledges funding from the H2020 EU project TYPES (grant no. 653449). R.C. acknowledges funding from H2020 EU project ReCRED (grant no. 653417).

## References

- [1] Facebook ads api. <https://developers.facebook.com/docs/graph-api>.
- [2] World bank human development index. <http://hdr.undp.org/en/content/human-development-index-hdi>.
- [3] World economic forum gender gap report. <http://reports.weforum.org/global-gender-gap-report-2016>.
- [4] Tim Berners-Lee. Long live the web. *Scientific American*, 303(6):80–85, 2010.
- [5] Ronald Brown, David Barram, and Larry Irving. *Falling through the net: A Survey of the "have nots" in rural and urban America*. United States Department of Commerce, 1995.
- [6] Moira Burke and Robert Kraut. Using facebook after losing a job: Differential benefits of strong and weak ties. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 1419–1430. ACM, 2013.
- [7] Samprit Chatterjee and Ali S Hadi. *Regression analysis by example*. John Wiley & Sons, 2015.
- [8] Benjamin M Compaine. *The digital divide: Facing a crisis or creating a myth?* Mit Press, 2001.
- [9] Jeff B Cromwell, Walter C Labys, and Michel Terraza. *Univariate tests for time series models*, volume 99. Sage, 1994.
- [10] Nicolas Friederici, Sanna Ojanperä, and Mark Graham. The impact of connectivity in africa: Grand visions and the mirage of inclusive digital development. *Electronic Journal of Information Systems in Developing Countries*, 79(2):1–20, 2017.
- [11] David Garcia, Ingmar Weber, and Venkata Rama Kiran Garimella. Gender asymmetries in reality and fiction: The bechdel test of social media. In *Eighth International AAAI Conference on Weblogs and Social Media*, 2014.

- [12] Laura K Gee, Jason J Jones, Christopher J Fariss, Moira Burke, and James H Fowler. The paradox of weak ties in 55 countries. *Journal of Economic Behavior & Organization*, 133:362–372, 2017.
- [13] World Bank Group. *World Development Report 2016: Digital Dividends*. Washington, D.C.: World Bank, 2016.
- [14] Eszter Hargittai and Yuli Patrick Hsieh. Digital inequality. pages 129–150. Oxford University Press, 2013.
- [15] James Hendler and Jennifer Golbeck. Metcalfe’s law, web 2.0, and the semantic web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(1):14–20, 2008.
- [16] Robert Jensen. The digital provide: Information (technology), market performance, and welfare in the south indian fisheries sector. *The quarterly journal of economics*, 122(3):879–924, 2007.
- [17] Manuel Koller and Werner A Stahel. Sharpening wald-type inference in robust regression for small samples. *Computational Statistics & Data Analysis*, 55(8):2504–2515, 2011.
- [18] Shirin Nilizadeh, Anne Groggel, Peter Lista, Srijita Das, Yong-Yeol Ahn, Apu Kapadia, and Fabio Rojas. Twitter’s glass ceiling: The effect of perceived gender on online visibility. In *ICWSM*, pages 289–298, 2016.
- [19] Pippa Norris. *Digital divide: Civic engagement, information poverty, and the Internet worldwide*. Cambridge University Press, 2001.
- [20] Martyn Plummer et al. Jags: A program for analysis of bayesian graphical models using gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing*, volume 124, page 125. Vienna, 2003.
- [21] Marian-Andrei Rizoiu, Lexing Xie, Tiberio Caetano, and Manuel Cebrian. Evolution of privacy loss in wikipedia. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining, WSDM ’16*, pages 215–224, New York, NY, USA, 2016. ACM.
- [22] Viktoria Spaiser, Shyam Ranganathan, Richard P Mann, and David JT Sumpter. The dynamics of democracy, development and cultural values. *PLoS one*, 9(6):e97856, 2014.
- [23] United Nations General Assembly. 60 resolution 252, 27 March 2006.
- [24] T Uteng. Gender and mobility in the developing world. *World Development Report*, pages 7778105–1299699968583, 2012.
- [25] Claudia Wagner, Eduardo Graells-Garrido, David Garcia, and Filippo Menczer. Women through the glass ceiling: gender asymmetries in wikipedia. *EPJ Data Science*, 5(1):5, 2016.

[26] World Wide Web Foundation. The web and rising global inequality, 2015.

# Supplementary Information

## Supplementary Table 1

| Gender | Facebook accounts | Median DAU  | Total population |
|--------|-------------------|-------------|------------------|
| Female | 646,953,680       | 375,148,380 | 2,422,856,560    |
| Male   | 800,026,950       | 391,673,365 | 2,485,459,305    |
| Both   | 1,446,980,630     | 766,821,745 | 4,908,315,865    |

Table 1: Total counts of Facebook accounts, DAU estimates, and total population covered by the dataset.

## Supplementary Text 1 - FGD as a function of other inequalities

Table 2 reports the Variance Inflation Factors of the variables in the FGD model. All factors are low enough to discard multi-collinearity in the linear model.

| Variable                       | VIF      |
|--------------------------------|----------|
| Education Gender Equality Rank | 2.079070 |
| Economic Gender Equality Rank  | 1.194107 |
| Health Gender Equality Rank    | 1.307413 |
| Political Gender Equality Rank | 1.239827 |
| Internet Penetration Rank      | 1.625328 |
| Income Inequality Rank         | 1.121453 |
| Total Population Rank          | 1.194382 |

Table 2: Variance Inflation Factors of independent variables in the FGD model.

Table 3 reports the detailed results of the FGD model fit and Table 4 reports the results of the same model when fitted with a robust regression method. Results are qualitatively similar, revealing that the FGD model result is robust to outliers.

| Term                      | Median estimate | 95% Credible Interval | p-value    |
|---------------------------|-----------------|-----------------------|------------|
| Intercept                 | <b>136.8</b>    | [121, 152.8]          | $p < 0.01$ |
| Education Equality Rank   | <b>-0.53</b>    | [-0.66, -0.40]        | $p < 0.01$ |
| Health Equality Rank      | <b>-0.27</b>    | [-0.37, -0.17]        | $p < 0.01$ |
| Economic Equality Rank    | <b>-0.16</b>    | [-0.25, -0.06]        | $p < 0.01$ |
| Political Equality Rank   | 0.05            | [-0.05, 0.14]         | 0.83       |
| Internet Penetration Rank | <b>-0.23</b>    | [-0.34, -0.12]        | $p < 0.01$ |
| Income Inequality Rank    | 0.01            | [-0.08, 0.11]         | 0.60       |
| Population Rank           | 0.009           | [-0.086, 0.10]        | 0.56       |
| $N$                       | 142             | $R^2$                 | 0.7411     |

Table 3: Regression results of FGD model. Estimates of p-values are based on the null hypothesis that the coefficient equals zero after 10,000 iterations.

Figure 5 shows the normal Q-Q plot and the histogram of residuals, which are distributed very close to normality. This is confirmed by a Shapiro-Wilk normality test, with a statistic of 0.99 and unable to reject the null hypothesis that residuals are normally distributed ( $p = 0.49$ ). Furthermore, residuals are uncorrelated with all gender equality variables (Near-zero Pearson correlation coefficients, with p-values above 0.9) and the square root of absolute residuals are not significantly correlated with predicted values.

To account for the possible role of GDP as a correlate of Internet penetration, we repeated the fit of the FGD model by replacing the Internet penetration rank

| Term                      | Estimate      | Standard Error | t-value | p-value        |
|---------------------------|---------------|----------------|---------|----------------|
| Intercept                 | <b>139.02</b> | 8.08           | 17.21   | $p < 10^{-10}$ |
| Education Equality Rank   | <b>-0.54</b>  | 0.06           | -8.35   | $p < 10^{-10}$ |
| Health Equality Rank      | <b>-0.29</b>  | 0.05           | -5.51   | $p < 10^{-6}$  |
| Economic Equality Rank    | <b>-0.13</b>  | 0.05           | -2.60   | 0.0104         |
| Political Equality Rank   | 0.06          | 0.05           | 1.18    | 0.2403         |
| Internet Penetration Rank | <b>-0.25</b>  | 0.06           | -4.32   | $p < 10^{-4}$  |
| Income Inequality Rank    | -0.001        | 0.05           | -0.02   | 0.98           |
| Population Rank           | -0.008        | 0.05           | -0.16   | 0.87           |

Table 4: Robust regression results of FGD model.

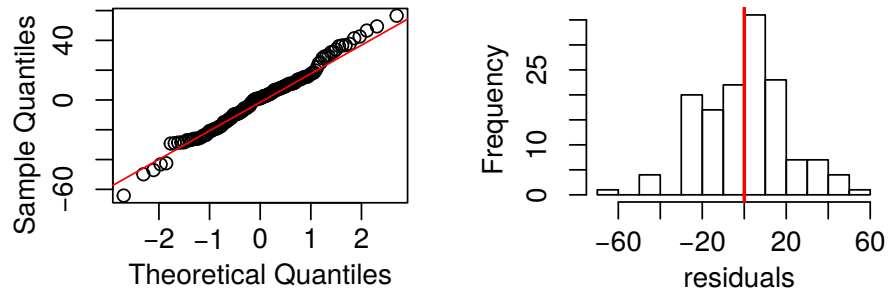


Figure 5: Left: Normal Q-Q plot of residuals of the FGD model. Right: Histogram of residuals. The regression residuals are close to normality.

with the GDP rank of the country. Results are reported on Table 5, revealing that the main result is robust to controlling for the wealth of countries.

| Term                    | Median estimate | 95% Credible Interval | p-value    |
|-------------------------|-----------------|-----------------------|------------|
| Intercept               | <b>147.3</b>    | [130.7, 163.7]        | $p < 0.01$ |
| Education Equality Rank | <b>-0.57</b>    | [-0.70, -0.44]        | $p < 0.01$ |
| Health Equality Rank    | <b>-0.25</b>    | [-0.35, -0.14]        | $p < 0.01$ |
| Economic Equality Rank  | <b>-0.17</b>    | [-0.26, -0.07]        | $p < 0.01$ |
| Political Equality Rank | 0.05            | [-0.05, 0.15]         | 0.80       |
| GDP Rank                | <b>-0.16</b>    | [-0.28, -0.04]        | 0.002      |
| Income Inequality Rank  | 0.02            | [-0.07, 0.12]         | 0.68       |
| Population Rank         | 0.01            | [-0.084, 0.114]       | 0.63       |
| $N$                     | 142             | $R^2$                 | 0.7228     |

Table 5: Regression results of FGD model with GDP instead of Internet penetration.

We repeated the fit of the FGD model for measurements of the DAU in twelve months between 2015 and 2016. Figure 6 shows the results of the fit for these alternative periods. The coefficient estimates barely depend on the period when the DAU are calculated and the  $R^2$  of the fits range between 0.726 and 0.757, confirming that our results are robust to fluctuations in the reporting of DAU through the Facebook API.

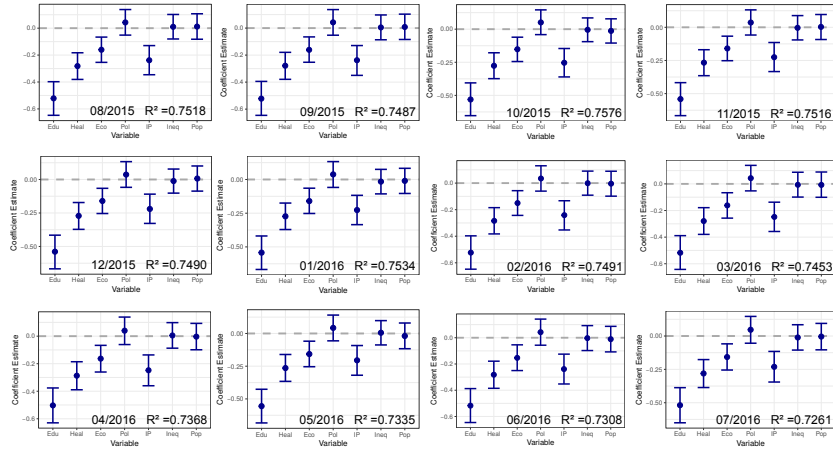


Figure 6: Results of repetition of the fit of the FGD model for 12 months between 2015 and 2016. The results are qualitatively stable across months.



## Supplementary Text 2 - Network effects

The results of the network effects model are shown on Figure 7. The model achieves a  $R^2 = 0.96$  on the logarithmic scale and a  $R^2 = 0.89$  on the linear scale of activity ratios per gender. Table 6 shows the detailed results of the model, evidencing the superlinear scaling ( $\alpha = 1.2$ ) and the difference between genders ( $\alpha_F = 0.25$ ).

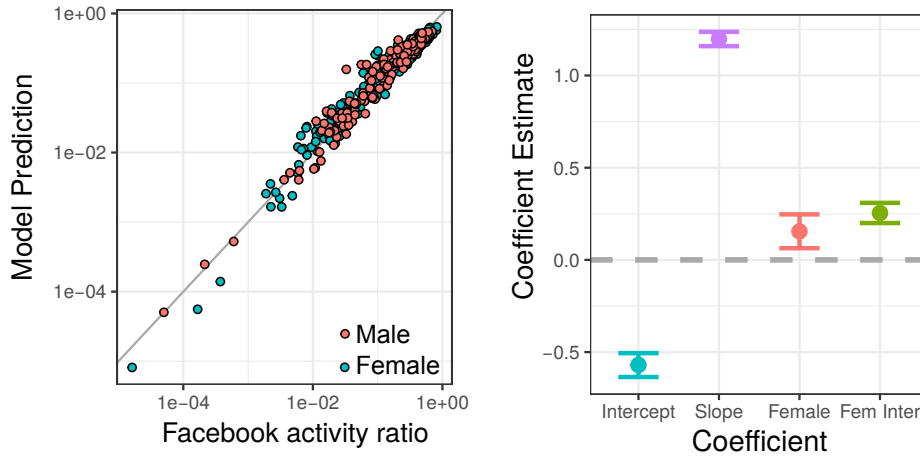


Figure 7: Results of network effects model fit. The left panel shows the comparison between empirical and predicted values, the right panel shows median estimates and 95% CI of model coefficients.

| Term       | Median estimate                | 95% Credible Interval | p-value    |
|------------|--------------------------------|-----------------------|------------|
| $\beta$    | <b>-0.57</b>                   | $[-0.63, -0.50]$      | $p < 0.01$ |
| $\alpha$   | <b>1.198</b>                   | $[1.16, 1.24]$        | $p < 0.01$ |
| $\beta_F$  | <b>0.15</b>                    | $[0.06, 0.25]$        | $p < 0.01$ |
| $\alpha_F$ | <b>0.25</b>                    | $[0.2, 0.31]$         | $p < 0.01$ |
| $N$        | 422 (211 countries, 2 genders) | $R^2$                 | 0.96       |

Table 6: Regression results of network effects model. Estimates of p-values are based on the null hypothesis that the coefficient equals one for  $\alpha$  and zero for the rest, after 10,000 iterations.

Figure 8 shows the analysis of the residuals of the model and the error in the linear scale of activity ratios per gender. Some small deviations from normality can be observed at the tails, corresponding to significant Shapiro-Wilk statistics of 0.94 and 0.95. Both types of residuals are uncorrelated with the Facebook presence ratio and do not appear to have a structure across predicted values. We identified some of the residual outliers, such as China and Tajikistan, which when removed do not have a qualitative impact in the results of the model fit and lead to residual distributions closer to normality.

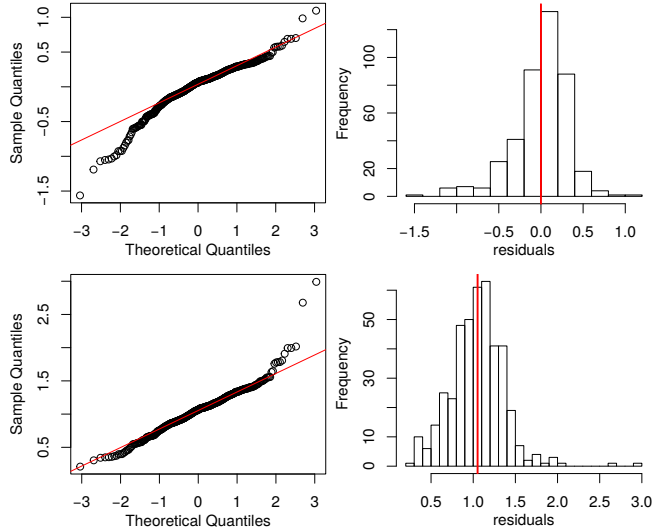


Figure 8: Analysis of residuals of network effects model. The top panels show the normal Q-Q plot and the histogram of residuals  $\phi$  of the model, and the lower panels the converse for multiplicative residuals in the linear scale of activity ratios per gender. Some minor deviations from normality can be observed in both.

We repeated the fit using a robust regression method, reporting the results on Table 7. While estimates slightly change, the qualitative results of a superlinear relationship that is stronger for female users still hold. This shows that our conclusions are robust to the influence of outliers.

| Term       | Estimate      | Standard Error | t-value | p-value        |
|------------|---------------|----------------|---------|----------------|
| $\beta$    | <b>-0.523</b> | 0.033          | -16.25  | $p < 10^{-10}$ |
| $\alpha$   | <b>1.19</b>   | 0.02           | 60.88   | $p < 10^{-10}$ |
| $\beta_F$  | <b>0.25</b>   | 0.05           | 5.43    | $p < 10^{-7}$  |
| $\alpha_F$ | <b>0.37</b>   | 0.03           | 12.59   | $p < 10^{-10}$ |

Table 7: Robust regression results of the Network effects model.

As with previous models, we evaluated the model of network effects over twelve months following our initial measurement. Figure 9 reports the overall results, showing no relevant decrease in  $R^2$  and generally the same result, where the parameter  $\alpha_F$  is significantly larger than zero and the parameter  $\alpha$  is significantly larger than one.

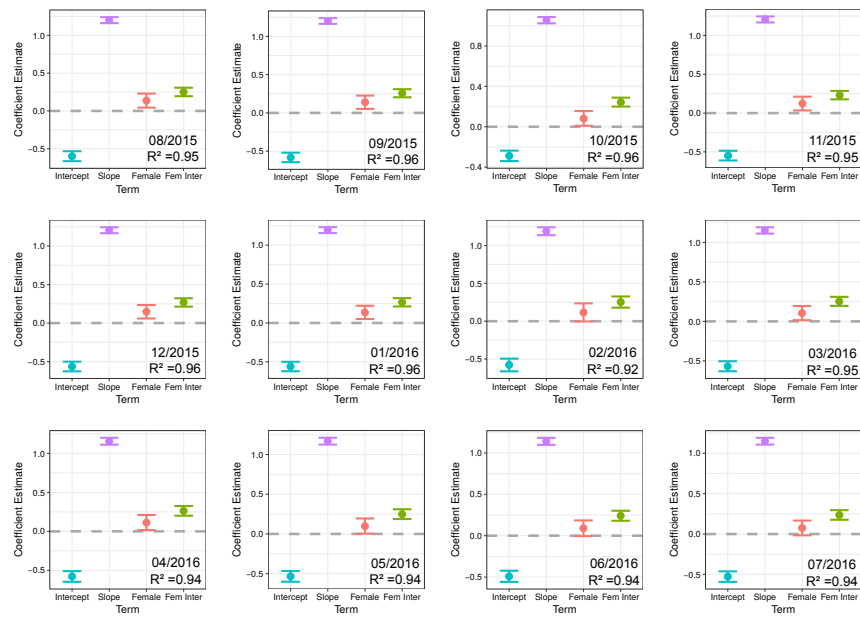


Figure 9: Results of repetition of the fit of the network effects model for 12 months between 2015 and 2016. The results are qualitatively stable across months.

### Supplementary Text 3 - Gender equality changes

Table 8 reports the Variance Inflation Factors of the variables in the model of economic gender equality changes. All factors are low enough to discard multi-collinearity in the linear model.

| Variable                      | VIF      |
|-------------------------------|----------|
| FGD Rank                      | 2.181729 |
| Economic Gender Equality 2015 | 1.209750 |
| GDP Rank                      | 2.007710 |
| Income Inequality Rank        | 1.100580 |
| Total Population Rank         | 1.059272 |

Table 8: Variance Inflation Factors of independent variables in the economic gender equality changes model.

Table 9 presents the detailed results of the model of changes in economic gender equality. Before fitting, we rescaled the ranked variables to have a value between zero and one to allow a better comparison of their relationships.

| Term                          | Median estimate | 95% Credible Interval | p-value    |
|-------------------------------|-----------------|-----------------------|------------|
| Intercept                     | -0.001          | [-0.03, 0.03]         | 0.57       |
| FGD Rank                      | <b>0.045</b>    | [0.067, 0.024]        | $p < 0.01$ |
| Economic Gender Equality 2015 | <b>-0.10</b>    | [-0.14, -0.06]        | $p < 0.01$ |
| GDP Rank                      | <b>0.064</b>    | [0.044, 0.083]        | $p < 0.01$ |
| Income Inequality Rank        | -0.003          | [-0.02, 0.013]        | 0.35       |
| Population Rank               | 0.013           | [-0.003, 0.03]        | 0.06       |
| $N$                           | 139             | $R^2$                 | 0.307      |

Table 9: Regression results of economic gender equality changes model. Estimates of p-values are based on the null hypothesis that the coefficient equals zero after 10,000 iterations.

The residuals of the model of changes in economic gender equality are distributed close to normality, as shown on Fig. 10, with a significant Shapiro-Wilk statistic of 0.97 and only some small deviations from normality at the tails. Residuals are uncorrelated with all independent variables and do not show signs of heteroscedasticity.

As with the other models, we repeated the fit of the model of changes in economic gender equality, using FGD measurements over 12 months. The results, shown on Figure 11, reveal that there is no qualitative change and that results are robust to any fluctuation of data reporting.

The results of a repetition of the fit with robust regression are shown on Table 10. The estimate of the coefficient for FGD is a bit smaller than in the MCMC fit (0.03 versus 0.045), but still significant and on the same direction. All results are qualitatively unchanged, showing that any outliers do not influence the conclusions of our analysis.

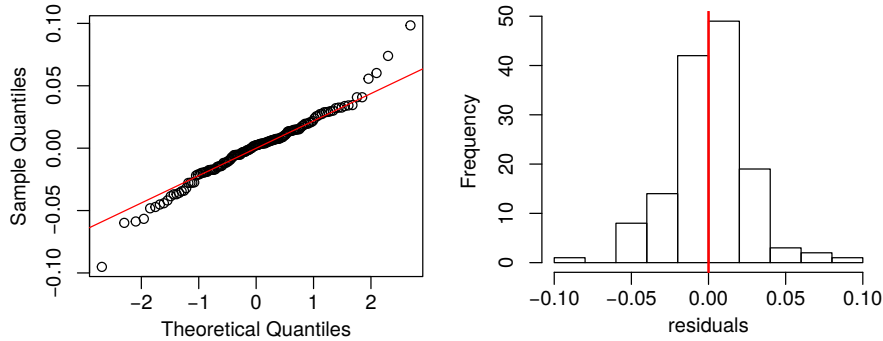


Figure 10: Analysis of residuals of the economic gender equality changes model. The left panel shows the normal Q-Q plot of residuals of the model, and the right panel their histogram. Some minor deviations from normality can be observed in both.

| Term                          | Estimate      | Standard Error | t-value | p-value       |
|-------------------------------|---------------|----------------|---------|---------------|
| Intercept                     | 0.05          | 0.015          | 0.33    | 0.74          |
| FGD Rank                      | <b>0.03</b>   | 0.01           | 2.548   | 0.012         |
| Economic Gender Equality 2015 | <b>-0.07</b>  | 0.02           | -3.1    | 0.003         |
| GDP Rank                      | <b>0.04</b>   | 0.01           | 4.4     | $p < 10^{-4}$ |
| Income Inequality Rank        | <b>-0.016</b> | 0.008          | -2      | 0.04          |
| Population Rank               | 0.015         | 0.008          | 1.99    | 0.05          |

Table 10: Robust regression results of economic gender equality changes model.

Since there is a significant and sizeable association between the ranks of FGD and the score of education gender equality (see Table 3), we need to test whether the negative coefficient of FGD in the model of economic equality changes can be attributed to the level of gender equality in education. We added a control term for the rank of gender education equality to the model and repeated the fit. We found a moderately higher  $R^2 = 0.31$  and no significant effect of education gender equality in the changes of economic gender equality. As shown on Figure 12, the rest of coefficients of the model remain qualitatively unchanged, evidencing that the relationship between FGD and changes in economic gender equality is not a confound with education gender equality.

We repeated the above model for the case of changes in education gender equality and for changes in political gender equality, including controls for the autocorrelation of each variable. We could not include an analysis of changes in health gender equality due to the fact that they do not change yearly, as that score depends on life expectancy and birth rates, two measurements that take many years to change. Figure 13 shows the coefficient estimates of both models, having  $R^2 = 0.17$  the model of education equality changes and  $R^2 = 0.02$  the model of political equality changes. The FGD has no significant effect in neither

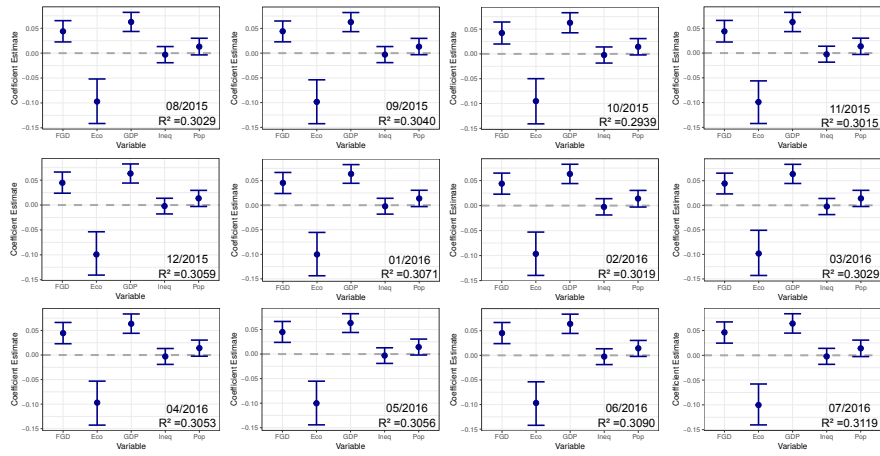


Figure 11: Results of repetition of the fit of the model of changes in economic gender equality for 12 months between 2015 and 2016. The results are qualitatively stable across months.

of these two equality metrics, indicating that the gender divide in social media only has a relationship to changes in economic gender equality.

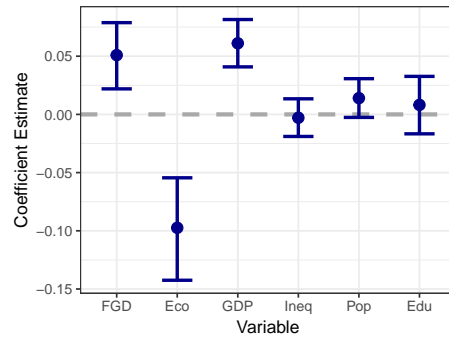


Figure 12: Repetition of the model of changes in economic gender equality, including a control for the rank of education gender equality in 2015 (Edu).

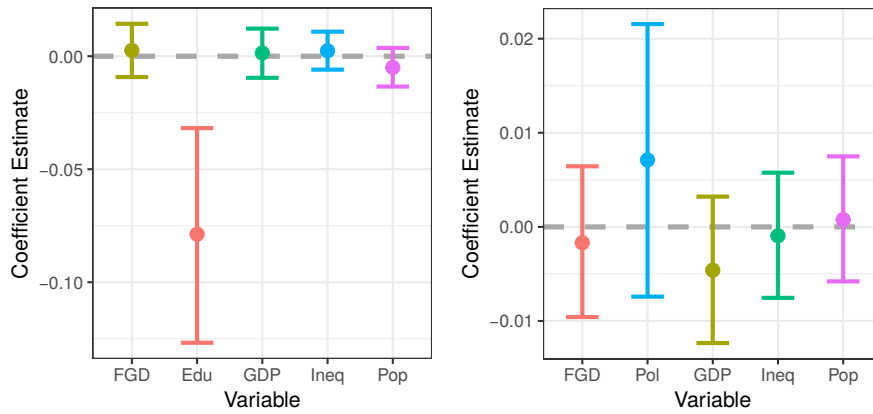


Figure 13: Model results for changes in education (left) and political (right) gender equality.