



# DEMOGRAPHIC RESEARCH

*A peer-reviewed, open-access journal of population sciences*

---

## ***DEMOGRAPHIC RESEARCH***

**VOLUME 43, ARTICLE 27, PAGES 779–816**

**PUBLISHED 8 SEPTEMBER 2020**

<https://www.demographic-research.org/Volumes/Vol43/27/>

DOI: 10.4054/DemRes.2020.43.27

*Research Article*

### **Monitoring global digital gender inequality using the online populations of Facebook and Google**

**Ridhi Kashyap**

**Reham Al Tamime**

**Masoomali Fatehkia**

**Ingmar Weber**

This publication is part of the Special Collection on Social Media and Demographic Research, organized by Guest Editor Emilio Zagheni.

© 2020 Kashyap, Fatehkia, Al Tamime, & Weber.

*This open-access work is published under the terms of the Creative Commons Attribution 3.0 Germany (CC BY 3.0 DE), which permits use, reproduction, and distribution in any medium, provided the original author(s) and source are given credit.*

*See <https://creativecommons.org/licenses/by/3.0/de/legalcode>.*

## Contents

1	Introduction	780
2	Background	783
2.1	Digital gender inequality	783
2.2	Monitoring development indicators with ‘big data’	785
3	Data	786
3.1	Online indicators: AdWords Gender Gap Index and Facebook Gender Gap Index	786
3.2	Dependent variable: Internet Use Gender Gap Index	791
3.3	Offline predictors	792
4	Methods	793
5	Results	795
5.1	Correlation analysis	795
5.2	Models predicting internet use gender gap	796
5.2.1	Offline model	796
5.2.2	Online model	798
5.2.3	Online–Offline model	799
5.3	Predicting gender gaps in digital skills	800
6	Discussion	804
7	Acknowledgements	807
	References	808
	Appendix A: Log transforming the gender gap indicators	814
	Appendix B: Comparing models with likelihood ratio test	816

## **Monitoring global digital gender inequality using the online populations of Facebook and Google**

**Ridhi Kashyap<sup>1</sup>**

**Masoomali Fatehkia<sup>2</sup>**

**Reham Al Tamime<sup>3</sup>**

**Ingmar Weber<sup>2</sup>**

### **Abstract**

#### **BACKGROUND**

In recognition of the empowering potential of digital technologies, gender equality in internet access and digital skills is an important target in the United Nations (UN) Sustainable Development Goals (SDGs). Gender-disaggregated data on internet use are limited, particularly in less developed countries.

#### **OBJECTIVE**

We leverage anonymous, aggregate data on the online populations of Google and Facebook users available from their advertising platforms to fill existing data gaps and measure global digital gender inequality.

#### **METHODS**

We generate indicators of country-level gender gaps on Google and Facebook. Using these online indicators independently and in combination with offline development indicators, we build regression models to predict gender gaps in internet use and digital skills computed using available survey data from the International Telecommunications Union (ITU).

#### **RESULTS**

We find that women are significantly underrepresented in the online populations of Google and Facebook in South Asia and sub-Saharan Africa. These platform-specific gender gaps are a strong predictor that women lack internet access and basic digital skills in these populations. Comparing platforms, we find Facebook gender gap indicators perform better than Google indicators at predicting ITU internet use and low-

---

<sup>1</sup> Department of Sociology, Leverhulme Centre for Demographic Science and Nuffield College, University of Oxford, UK. Address correspondence to: [ridhi.kashyap@nuffield.ox.ac.uk](mailto:ridhi.kashyap@nuffield.ox.ac.uk).

<sup>2</sup> Qatar Computing Research Institute, Hamad Bin Khalifa University, Qatar.

<sup>3</sup> University of Southampton, UK.

level digital-skill gender gaps. Models using these online indicators outperform those using only offline development indicators. The best performing models, however, are those that combine Facebook and Google online indicators with a country's development indicators such as the Human Development Index.

## **CONTRIBUTION**

Our work highlights how appropriate regression models built on novel, digital data from online populations can be used to complement traditional data sources to monitor global development indicators linked to digital gender inequality.

## **1. Introduction**

The internet has revolutionized how individuals seek information, communicate, and access goods and services. By lowering the cost of information and connectivity, the internet has tremendous potential to help meet sustainable development goals (SDGs) and this role is acknowledged in different United Nations' (UN) SDG targets (United Nations 2016). Digital literacy is recognized as part of the right to education (Goal 4). The commitment to ensuring equitable access to the internet and other information and communication technologies (ICTs) is noted as part of the SDG on gender equality (Goal 5), which pledges to “enhance the use of [...] information and communication technology to promote the empowerment of women.”<sup>4</sup>

Even as internet access has proliferated, ‘digital divides’ or inequalities in access to and use of the internet persist (Scheerder, van Deursen, and van Dijk 2017; Robinson et al. 2015). These inequalities are an increasingly important dimension of population inequality as digitalization continues to unfold and affect domains of social, demographic, and economic life (DiMaggio et al. 2004; Robinson et al. 2015; Hjort and Poulsen 2019; Billari, Giuntella, and Stella 2019; Rotondi et al. 2020). Online inequalities often mirror sociodemographic, offline inequalities, and the digital divide by gender is one widely noted dimension of this inequality. According to the UN’s specialized agency for ICTs, the International Telecommunication Union (ITU), over 250 million fewer women are online than men and gender gaps in internet use tend to be greater in developing countries (International Telecommunication Union 2017). Digital gender inequality has significant implications for social and demographic outcomes. Research from both developed and developing countries has shown that internet and mobile phone use among women positively affects outcomes linked to their reproductive health and their children’s health, as well as bolsters their decision-making

---

<sup>4</sup> <https://unstats.un.org/sdgs/metadata/>

power and economic prospects (Lagan, Sinclair, and Kernohan 2011; Lund et al. 2014a, 2014b; Beck et al. 2014; Suri and Jack 2016; Dettling 2017; Rotondi et al. 2020). The payoffs from technology adoption for these social and demographic outcomes are higher among the least developed countries (Rotondi et al. 2020).

The increasing recognition of the role of digital technologies in the empowerment of women and the persistence of digital gender inequality has led several UN agencies such as the ITU and UNESCO to endorse targets calling for gender equality in internet use and access to broadband (Broadband Commission 2013; International Telecommunication Union 2015; European Parliament 2018). However, the lack of gender-disaggregated data on internet use remains “one of the key barriers” in monitoring progress towards these development targets (Broadband Commission 2013: 19). Official, nationally representative, gender-disaggregated statistics on internet use lack regular production, and data availability on these indicators is especially limited in developing countries (Hafkin and Huyer 2007). The most recent release of the ITU’s World Telecommunications Indicators Database provides data on internet use by gender for 83 countries, with country coverage being especially sparse for low- and lower-middle-income countries where data are only available for 16 countries (out of 64) (International Telecommunication Union 2018).<sup>5</sup> In the same database, data on gender gaps in specific digital skills are available for even fewer countries than data on more general internet use measures, with coverage of under 50 countries for all indicators. Routine survey data collection on individual-level ICT use within households is expensive, and while some population censuses are able to collect information on internet or mobile availability at the household level, intra-household inequalities are not captured in these data sources (Fatehkia, Kashyap, and Weber 2018; Feehan and Cobb 2019).

Given existing data gaps on digital gender inequalities, particularly in the context of less developed countries, previous work by Fatehkia, Kashyap, and Weber (2018) has highlighted how digital trace data on the gender composition of Facebook users available from its advertising platform has the potential to help fill these data gaps in measuring global gender disparities in internet use. This paper builds on the aforementioned study by examining the potential of another novel data source to measure global digital gender gaps: Google’s advertisement impression estimates (AdWords). Whereas Facebook reaches a mere 60% of internet users<sup>6</sup>, according to Google’s own claims “the Google Display Network reaches 90% of internet users worldwide.”<sup>7</sup> Similarly to Facebook, Google allows advertisers to estimate the reach of

---

<sup>5</sup> The latest release of the data covers surveys that were fielded in ITU member states between 2013 and 2017.

<sup>6</sup> 2.2B out of 3.6B – see <https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/> and <https://www.statista.com/statistics/273018/number-of-internet-users-worldwide/>.

<sup>7</sup> <https://ads.google.com/home/how-it-works/display-ads/>.

their campaigns by showing them an estimate of the expected number of weekly impressions, i.e., the number of times an ad is expected to be shown on a search result page or another site on the Google Display Network.<sup>8</sup> These estimates can be filtered based on demographic targeting criteria such as age and gender, and are available for over 200 countries.<sup>9</sup> Our paper examines AdWords' potential for predicting gender gaps in internet use and compares the performance of AdWords and Facebook, both independently and together. Furthermore, we extend previous work by assessing the kinds of digital-skill gender gaps the AdWords and Facebook indicators can serve as proxies for.

We generate a country-level dataset combining (1) online indicators on gender gaps derived from Google AdWords impressions and Facebook audience estimates, (2) the latest available statistics collected using surveys on gender gaps in internet use and different dimensions of digital skills available from the ITU, and (3) offline indicators related to a country's overall level of development (e.g., Human Development Index (HDI)) and gender gaps (e.g., education, occupations). With this dataset, we estimate models to predict ITU estimates of gender gaps in internet use and different digital skills using both online indicators and a combination of online and offline indicators.

Our results show that Facebook and Google online indicators are both highly correlated with ITU internet use gender gaps, as well as low-level digital skills such as using copy and paste tools, transferring files, and sending emails. Even though not all internet users are Facebook or Google users, these correlations suggest that when women are missing from the populations of these two online platforms it is a strong indicator that women are not online and lack digital skills in the populations of their countries. Although regression models that use Facebook online indicators show better predictive performance than Google AdWords, models that combine Facebook and Google online indicators with a country's offline development indicators offer the best predictive performance for internet use gender gaps. Together with the HDI, the Facebook and AdWords gender gap indicators are able to explain about 80% of the variation in global internet gender gaps. Facebook indicators are better able to predict low-level digital skills than AdWords indicators.

Our approach demonstrates how anonymous aggregate data from two of the biggest online populations can be repurposed to measure important sustainable development indicators linked to digital gender equality. This approach comes with big gains in geographical coverage for less developed countries where existing survey data are often lacking or infrequently collected. We also show that women are significantly underrepresented on two of the largest online platforms, Google and Facebook, in sub-Saharan Africa and South Asia. This is a crucial bias for researchers to bear in mind

---

<sup>8</sup> <https://support.google.com/google-ads/answer/2404190?hl=en>.

<sup>9</sup> <https://support.google.com/google-ads/answer/6320>.

when using these digital data sources for social and demographic measurement, as well as for development practitioners when using these platforms for media outreach and informational campaigns.

## **2. Background**

### **2.1 Digital gender inequality**

Sociodemographic, economic, and cultural factors have been shown to be associated with gender gaps in internet access and use (Scheerder, van Deursen, and van Dijk 2017). Evidence gathered from developed countries indicates that the gap between men and women has closed in terms of internet access as internet penetration has increased. However, a ‘second-order’ digital divide in terms of patterns of use and skills has been found, with women showing lower frequency of use, a narrower range of online activities, and lower likelihood of reporting strong internet skills, even if their actual web use skills are not lower than those of men (Robinson et al. 2015; Ono and Zavodny 2007; Hargittai and Shafer 2006; DiMaggio et al. 2004). These findings have led scholars to suggest that the digital divide is better conceptualized in terms of a spectrum of skills instead of a binary classification of whether an individual is an internet user or not (Hargittai 2002). Digital skills are important for understanding whether the benefits of the internet are accrued evenly or whether digital inequality is further exacerbated as internet penetration increases (Hargittai and Shafer 2006).

While literature on digital gender inequality in the context of developing countries is comparatively limited, research suggests that gender gaps in internet access exist across different age groups (Antonio and Tuffley 2014). Internet access gender gaps in developing countries reflect broader structural inequalities that women face in terms of access to education, employment, and income (WWW Foundation 2015; Robinson et al. 2015; Hilbert 2011; Hafkin and Huyer 2007). In addition to these socioeconomic barriers to access, studies have also documented how cultural norms in patriarchal contexts may also impede women’s internet use, particularly when internet access is mediated via men (Abu-Shanab and Al-Jamal 2015; WWW Foundation 2015; Gurumurthy and Chami 2014; Intel 2012). The discussion of the digital divide in the context of developing countries has focused largely on access or general use inequalities, as survey data on digital skills or digital literacy in these country contexts are especially limited (Hafkin and Huyer 2007; Broadband Commission 2013; WWW Foundation 2015).

Digital gender inequalities have significant implications for demographic and social outcomes linked to population health, well-being, and social and economic

empowerment. Social scientists have noted that the internet and mobile phones can promote access to information, boost connectivity and social learning, and provide cost-efficient access to vital services (DiMaggio et al. 2004; Unwin and Unwin 2009; Aker and Mbiti 2010; Suri and Jack 2016; Hjort and Poulsen 2019). Empirical findings provide support for the positive effects of these technologies, and indicate that these effects are especially salient for women. For example, in the German context, Dettling (2017) finds that the diffusion of the internet promotes labor force participation among married women and Billari, Giuntella, and Stella (2019) find that the internet enables women to better reconcile work and motherhood, thereby boosting well-being and (wanted) fertility. Conversely, in a high-fertility setting in Malawi, mobile phone ownership among women is associated with smaller ideal family size, lower parity, and increased child spacing (Billari, Rotondi, and Trinitapoli 2020). Work by Beck et al. (2014) and Lagan, Sinclair, and Kernohan (2011) suggests that women are more likely to draw on the internet to empower themselves regarding health-related information.

In the context of developing countries, where women often face greater barriers to accessing information, communication, and wider social networks, the payoffs of internet access – which has become cheaper and more widely available due to the rapid diffusion of mobile phones – are likely to be even greater. Rotondi et al. (2020) provide empirical support for this idea and find that mobile-phone diffusion over time within countries is associated with lower gender inequality, higher contraceptive uptake, and lower maternal and child mortality, and these associations are larger in least developed countries, even after controlling for other developmental processes such as economic growth and educational expansion. The same study further shows that mobile phone ownership among women in particular is linked to improvements in contraceptive knowledge and uptake, antenatal health behaviors, and greater equality in household decision-making in sub-Saharan Africa.

Despite the potential of digital technologies for bolstering key outcomes linked to demographic and social development, if digital gender inequality persists the positive benefits of technology cannot be fully realized. In light of this, digital gender inequalities are an important indicator to track in the context of the sustainable development goals, and this target is expressly recognized in Sustainable Development Goal 5 on gender equality. Nevertheless, there are significant data gaps that limit the regular monitoring of digital gender inequality, especially in less developed countries. In the latest edition of the ITU Telecommunications Indicators Database, survey data on internet use by gender were available for 83 countries, of which 41 were high-income countries, 26 were upper-middle income countries, 13 were lower-middle income, and 3 were low-income countries.<sup>10</sup> These data were collected from surveys that were fielded by ITU member states in the period 2013–2017. Coverage on digital skills indicators is

---

<sup>10</sup> Following the World Bank income classification of countries.



even more limited. For example, data on the basic digital skill of being able to transfer files are available for 47 countries, with 5 being low- and lower-middle-income countries. These data gaps show the need to consider alternative data sources.

## **2.2 Monitoring development indicators with ‘big data’**

In recent years, several researchers and international organizations have begun to explore the potential for ‘big data’ sources to overcome challenges associated with limited data coverage on development indicators, particularly in developing countries (Weber, Kashyap, and Zagheni 2018; di Bella, Leporatti, and Maggino 2016; IUSSP 2015; IEAG 2014; Letouze and Jutting 2014). This work has used diverse big data sources, ranging from the use of mobile call log data to predict income in African countries (Blumenstock, Cadamuro, and On 2015; Mao et al. 2015), mobile transport apps to predict social disadvantage (Benita 2019; Tan, Zhao, and Huang 2019), night-time satellite data to measure poverty (Elvidge et al. 2009), and web search and public social media posts to predict unemployment and health outcomes (Resce and Maynard 2018; Nuti et al. 2014; Choi and Varian 2012). Despite weaknesses in big data sources, such as issues of non-representativeness and limited metadata to understand the data-generating process, a strength of these data sources is their higher frequency or real-time measurement, which make them promising for ‘nowcasting’ (Salganik 2017; di Bella, Leporatti, and Maggino 2016). Nowcasting is typically employed when the actual value of the indicator of interest will only be known with a significant delay, creating the need to “predict the present” (Choi and Varian 2012). A further strength of some of these data sources, particularly those that relate to large online platforms, is their wider geographical coverage.

One of the big data sources used in this study, Facebook’s advertisement audience estimates available from the platform’s marketing application programming interface (API), can be queried for information on the number of Facebook users by various demographic characteristics and can be thought of as a kind of real-time census of the platform’s user base. These data have been leveraged to study population health (Araújo et al. 2017; Chunara et al. 2013), to provide demographic estimates of migration (Zagheni, Weber, and Gummadi 2017) and male fertility (Rampazzo et al. 2018), and to generate gender inequality measures, including, most relevantly for this study, country-level internet gender gaps (Fatehkia, Kashyap, and Weber 2018). Even though not all Facebook users are internet users, Fatehkia, Kashyap, and Weber (2018) highlight how gender gaps in Facebook populations serve as a robust predictor of internet and mobile access gender gaps. Gender gaps in Facebook use across countries have also been shown to be correlated with different domains of gender inequality more

generally, including education, health, and economic opportunity. In countries where gender inequality in socioeconomic domains is larger, men also outnumber women on Facebook (Garcia et al. 2018). These patterns have also been found at the sub-national level in diverse countries such as India, where significant sub-national variation in gender gaps in Facebook use exist, and gender gaps tend to be larger in states with greater gender inequality in schooling and lower levels of economic and social development (Mejova et al. 2018).

Our work applies the approach developed in Fatehkia, Kashyap, and Weber (2018), which found an indicator called the ‘Facebook Gender Gap Index’ to be highly correlated with ITU statistics on internet gender gaps collected using surveys fielded by national statistical agencies. That study further found that the predictive performance of the Facebook indicator was enhanced when combined with offline indicators linked to a country’s development. We build on that work by (1) going beyond internet-use gender gaps to also explore correlations with different digital skill gender gaps, and by (2) including data from Google, the biggest online advertising platform, to explore its potential as a complementary online data source to monitor internet access gender gaps. To the best of our knowledge, ad impression estimates for Google’s Adwords platform have not yet been used to monitor digital gender gaps or any other targets related to the SDGs.

### **3. Data**

Our dataset comprises (1) online indicators derived from advertisement impression estimates available from Google AdWords and advertisement audience estimates from Facebook’s Marketing API, (2) an extensive range of offline indicators related to a country’s level of development (e.g., GDP per capita, HDI) and gender inequalities (e.g., gender gaps in literacy), and (3) ITU data on internet use as well as specific ICT or digital skills by gender and country of the user, collected using nationally representative surveys, which we use to derive our ground truth internet use or digital skills gender gap indicators. These different indicators are described in this section. Table 1 provides summary statistics of the variables used in the analysis.

#### **3.1 Online indicators: AdWords Gender Gap Index and Facebook Gender Gap Index**

The online data we use come from publicly accessible, anonymous, and aggregate data that online platforms provide to advertisers to estimate the potential reach or audience

size of their advertisement campaigns. Facebook's advertisement audience estimates used in Fatehkia, Kashyap, and Weber (2018) are the number of monthly active Facebook users (MAUs) disaggregated by various geographic and demographic attributes, such as user age and gender. By contrast, the AdWords data do not provide information on the number of users but instead the number of impressions. Impressions are counted each time an ad is shown on a search result page or other site on the Google Network. The provided impression counts are weekly estimates and can be disaggregated by various geographic and demographic attributes. Google allows advertisers to create different kinds of campaigns on AdWords, including search network only, display network only, and video, shopping, and universal apps. We selected the display network campaign in order to retrieve the number of impressions for this study. The Google Display Network consists of over two million websites, videos, and apps where Google ads can appear.<sup>11</sup>

AdWords' reach estimates have the potential to be used for social science research questions pertaining to the attributes of the world's population, as they capture information related to the world's online population. Even though the use of AdWords' impressions for social science applications has been limited, Facebook's advertisement-audience estimates have been used in a similar vein as a type of 'digital census' (Zagheni, Weber, and Gummadi 2017; Fatehkia, Kashyap, and Weber 2018). Zagheni, Weber, and Gummadi (2017) find that the Facebook data can provide good demographic estimates of quantities such as percentages of the population of a particular nationality. As Google claims that its "Google Display Network reaches 90% of internet users worldwide"<sup>12</sup> – more than Facebook – our study explores the use of AdWords' number of impressions as a novel data source to monitor digital gender gaps.<sup>12</sup> An important difference to note is that compared to Facebook's user estimates, in AdWords more active users are likely to cause more impressions, potentially leading to certain biases. For the purpose of this study, the number of impressions was disaggregated by age and gender and was collected in June 2018 from 200 countries.<sup>13</sup> Also, we selected the language to include all languages spoken in every country. The 'unknown' number of impressions was excluded from the analysis as it related to audiences whose age and gender were not identified. The percentage of impressions with unknown gender varies across countries; for example, 42.4% of impressions in Switzerland and 20% of impressions in Afghanistan were of unknown gender. The number of impressions was used to compute the AdWords Gender Gap Index (ADW GGI) for each country as follows:

---

<sup>11</sup> <https://support.google.com/google-ads/answer/117120>.

<sup>12</sup> Information and documentation about AdWords is available here: [https://en.wikipedia.org/wiki/Google\\_Ads](https://en.wikipedia.org/wiki/Google_Ads) and <https://support.google.com/google-ads/answer/6319>.

<sup>13</sup> <https://developers.google.com/adwords/api/docs/appendix/geotargeting>

$$\text{ADW GGI} = \frac{\text{Female to male gender ratio of impressions}}{\text{Female to male gender ratio of the population}} \quad (1).$$

Following Fatehkia, Kashyap, and Weber (2018) and in the interest of measuring digital inequality rather than demographic gender imbalances in a population, we divide the gender ratio (female to male) of impressions on AdWords with the population gender ratio of the same age category as the AdWords ratio. For example, countries such as the United Arab Emirates (UAE) have a much larger male than female population due to the influx of foreign male workers. Correspondingly, observing a gender imbalance in terms of the number of Google impressions (or users) for the UAE could be merely reflecting the (offline) population gender imbalance. We obtained the population gender ratios for various age groups from the UN World Population Prospects Database (United Nations Population Division 2017).

Similarly, Facebook advertisement audience estimates were collected in May 2018 for 193 countries from Facebook’s Marketing API.<sup>14</sup> The data were on monthly active Facebook users disaggregated by age, gender, and country. The Facebook Gender Gap Index (FB GGI) for each country can be defined as:

$$\text{FB GGI} = \frac{\text{Female to male gender ratio of users on Facebook}}{\text{Female to male gender ratio of the population}} \quad (2).$$

We have excluded some countries from the analysis because Google AdWords provides vague estimates for those countries, such as “> 1B”, which do not permit us to calculate ratios. Countries affected by these vague estimates include India, Japan, Turkey, and the United States. This issue does not affect the Facebook data. Furthermore, some countries such as “Saint Pierre and Miquelon” were excluded because they are listed under AdWords as countries but are not recognized by the World Bank or the UN. From the Facebook data we excluded countries with less than one million users. After imposing these restrictions, we were left with data for 176 countries from Facebook and 166 countries from AdWords for the age group 18+. Both Google and Facebook support further filtering of impressions/users with certain characteristics. In particular, we compute variants of ADW GGI and FB GGI for different age groups of users (e.g., 25–29 year olds). Summary statistics for FB and ADW GGI measures are presented in Table 1.

All gender gap indicators used in this study take values greater than zero, with values less than 1.0 indicating countries where there is a gender gap that disadvantages women. Values close to 1.0 are desirable as they indicate gender parity and countries

---

<sup>14</sup> Information and documentation about Facebook’s Marketing API is available here: <https://developers.facebook.com/docs/marketing-apis>.

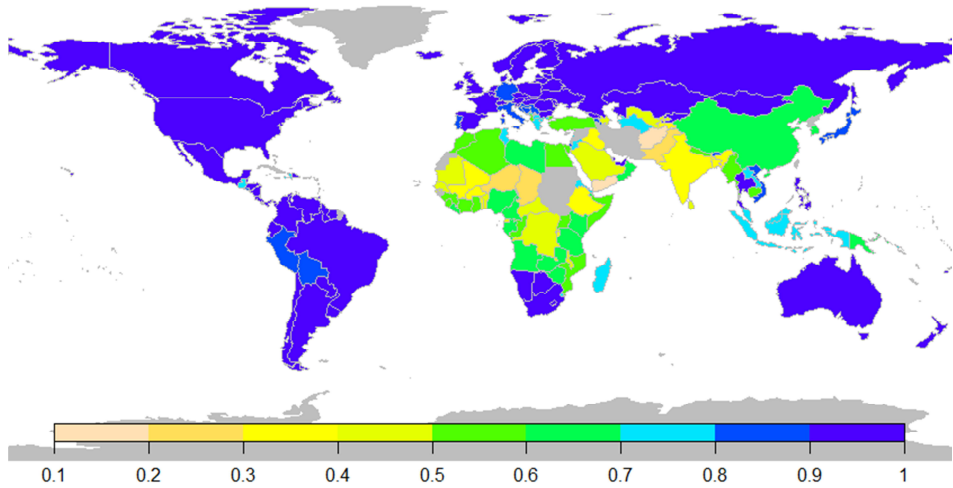
where the gender gap has closed. We truncated values larger than 1.0, corresponding to women doing better than men, as our focus is on gender equality rather than women's empowerment. Consequently, our measure does not differentiate between countries that have attained parity (a gender gap index equal to 1.0) and those where women have surpassed men (a gender gap index greater than 1.0). This is in line with the methodology of the Global Gender Gap Report (GGGR) published by the World Economic Forum (World Economic Forum 2016) and follows the approach used by Fatehikia, Kashyap, and Weber (2018), which helps facilitate a comparison of our results. Nevertheless, in the Appendix we also present results of our analysis where the gender gap indicators are uncapped and also log-transformed to correct for skewedness.

Figure 1 depicts the FB GGI (18+) (panel a) and the ADW GGI (18+) (panel b) for all countries for which we are able to compute them. Countries in Africa and South Asia show the largest gender gaps, followed by those in North Africa and the Middle East. Southern Africa is an exception in Africa, where both FB and ADW indicators show similar levels of gender parity as seen in South and North America and most parts of Western Europe. Within Europe, Southern Europe has larger gender gaps disfavoring women than Northern and Western Europe. The colors in both the upper and lower panels in Figure 1 largely coincide. This is further confirmed in Figure 2, which shows that the FB GGI and ADW GGI for countries are strongly correlated, with a Pearson's correlation coefficient of 0.85 and a Kendall rank correlation coefficient of 0.66. In Figure 2 there are on average a greater number of points below the  $x = y$  diagonal than above it. In about 63% of the cases,  $FB\ GGI > ADW\ GGI$ , and in 26% of the cases  $ADW\ GGI > FB\ GGI$ , with 11% being equal.

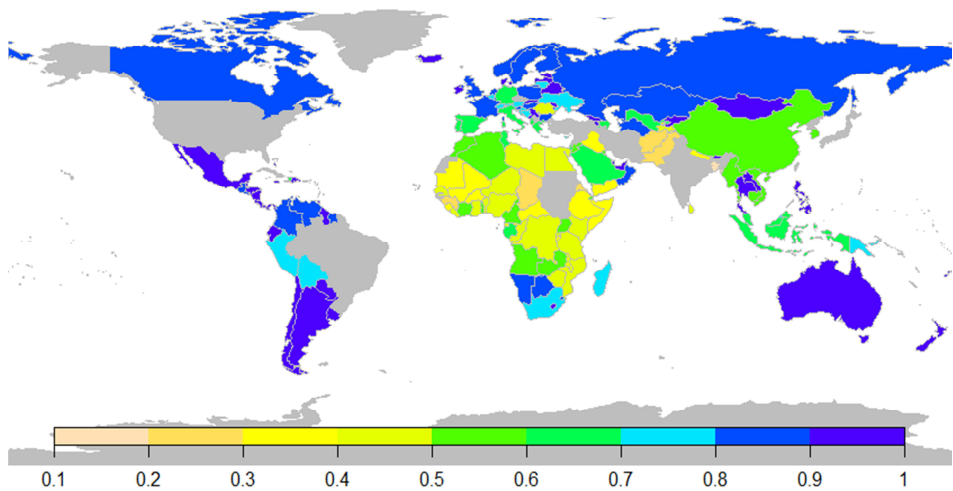
However, cases with one value that is systematically different from the other do not seem to follow any distinctive regional pattern. This suggests that the same countries tend to score higher on the FB GGI, indicative of greater gender equality, than on the ADW GGI. Although this could potentially indicate differential patterns of use of Google and Facebook by gender, this could also reflect the fact that the ADW GGI is based on impressions whereas the FB GGI is based on users. If men are more active users of the internet and generate a greater number of impressions or page views on Google, then this is likely to result in greater gender gaps in the ADW measure than in the FB measure.

**Figure 1: World maps showing Facebook (panel a) and AdWords (panel b) gender gap indices**

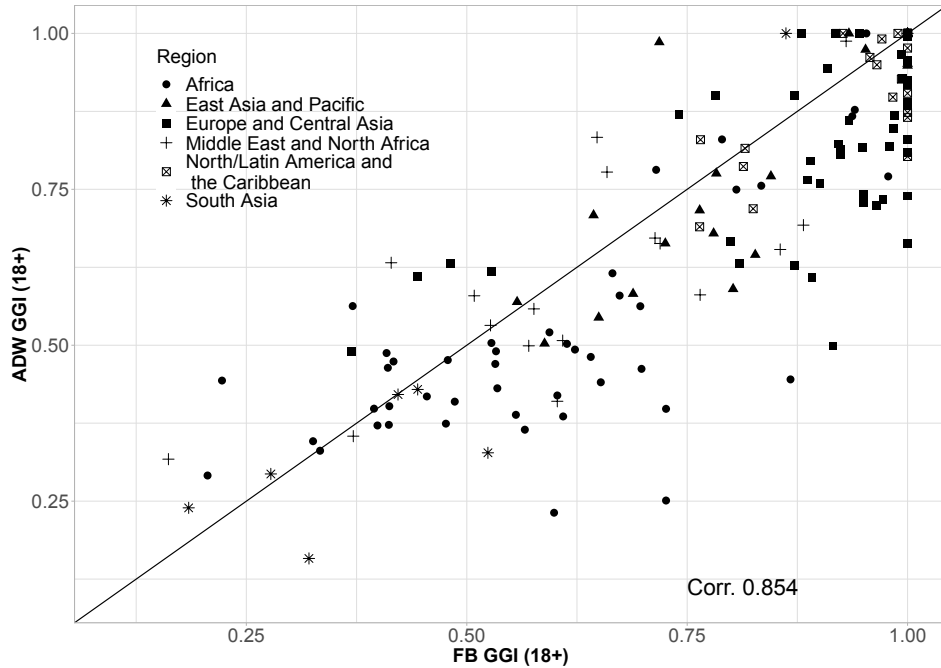
FB GGI (ages 18+)



ADW GGI (ages 18+)



**Figure 2: The FB GGI (ages 18+) against the ADW GGI (ages 18+). Each point indicates a country with shape coded by world region. The line is the  $x = y$  diagonal.**



### 3.2 Dependent variable: Internet Use Gender Gap Index

Although both Google and Facebook have large numbers of users, not all internet users use these platforms. Thus, to evaluate to what extent a gender gap index computed using the audiences of these platforms captures internet-use gender gaps as well as gender gaps in other digital skills, we need to correct our online indicators for potential biases. To do this, we need to compare and validate the online indicators against ground truth measures of internet use by gender and location. We use data on internet use by gender of the user and country reported in the ITU World Telecommunications Indicators Database as this ground truth measure (International Telecommunication Union 2018).

The ITU data provide the proportion of individuals using the internet by gender to give gender-specific internet penetration rates (e.g., 40% of women in a given country

are internet users). The data are collected using nationally representative surveys fielded by national statistical agencies in the ITU's member states. The latest release of the data covers surveys that were fielded in member states between 2013 and 2017. The data are thus available for different years for different countries based on whenever the survey was fielded there. The Internet Use Gender Gap Index (Internet GGI) for a country using ITU data is defined as:

$$\text{Internet GGI} = \frac{\% \text{ of female population using internet}}{\% \text{ of male population using internet}} \quad (3).$$

In addition to data on internet use by gender, the ITU also provides data on different types of digital skills by gender (International Telecommunication Union 2018). These measures capture a range of specific skills, such as being able to copy or move a file or folder, sending e-mails with attached files, and using basic arithmetic formula in a spreadsheet (see Table 1). These measures are collected via a survey, and are self-reported in response to questions that ask whether the user has undertaken the specific computer-related activity in the last three months (International Telecommunication Union 2016). The skills data are available for considerably fewer countries than internet use by gender (see Tables 2 and 3). Where data are available, we can also compute a digital skill-specific GGI (DS GGI) akin to the Internet GGI. However, due to a sparsity of ground truth for training purposes, our ability to use all of these measures as dependent variables in predictive models is limited.

### 3.3 Offline predictors

Our dataset includes offline predictors that have been associated with internet use gender gaps (see Appendix B of Fatehkia, Kashyap, and Weber 2018 for a full list). These include factors associated with a country's overall level of development, captured in indicators such as the GDP per capita and different dimensions of the HDI; its ICT infrastructure, captured in indicators such as the internet penetration rate; and gender-specific development indicators such as gender gaps in educational attainment or gender gaps in senior managerial work. In addition to these economic and development indicators used in previous work, we include six cultural variables, drawing on Hostede's cultural dimensions theory that classifies different countries based on the six dimensions of power distance, uncertainty avoidance, individualism/collectivism, masculinity/femininity, long/short-term orientation, and indulgence/restraint (Hofstede 2011).



## **4. Methods**

We fitted three types of ordinary least squares (OLS) regression models to predict the internet GGI, our ground truth measure of internet-use gender gaps. These are: (1) online models using different ADW GGI and FB GGI predictor variables, (2) online–offline models to assess if some of the biases from using only online indicators for measuring internet-use gender gaps can be corrected when combined with offline variables, and (3) an offline model using only offline predictors. We also attempted to use FB and ADW GGI variables from different age groups in our online models. For models that relied on multiple variables, we performed variable selection using a greedy step-wise forward approach whereby variables were iteratively added to the model, starting with a model with just an intercept in order to increase the adjusted R-squared of the resulting model. Table 1 presents the descriptive statistics for the outcome variables used and the online indicators and offline predictors that were picked up by the step-wise models. This methodology was chosen to ensure comparability of the results with prior work (Fatehkia, Kashyap, and Weber 2018). The offline models are a benchmark against which we compare the predictions of the online models, which rely on data from Google and Facebook to generate insights about internet-use gender gaps. Due to the smaller number of countries for which ITU ground truth data are available for the digital skills GGI, when predicting these outcomes we estimated the best-fit two-variable models using either online or offline indicators. To evaluate the performance of different models in predicting the internet and DS GGI, we report three measures of model fit: (1) Adjusted R-squared, (2) Mean Absolute Error, and (3) Symmetric Mean Absolute Percentage Error (SMAPE). For the first one, larger values and values closer to 1.0 indicate better performance. For the last two, lower values and those closer to 0.0 indicate better performance. The SMAPE is computed using a Leave-One-Out cross-validation procedure where the model is fitted on all of the data except for one country, and the fitted model is then used to predict the left-out data point. The SMAPE provides a measure of out-of-sample performance on unseen data.

**Table 1: Summary statistics of the various online and offline variables used in the analysis**

Variable	Q1	Median	Mean	Q3	Std. dev.
<i>Dependent variables: Internet GGI</i>					
ITU Internet GGI	0.917	0.971	0.934	0.990	0.107
<i>Dependent variables: DS GGI</i>					
Using copy and paste tools	0.884	0.937	0.903	0.991	0.137
Using basic arithmetic formula in spreadsheet	0.794	0.865	0.860	0.979	0.131
Writing computer program using prog. lang.	0.321	0.426	0.472	0.568	0.206
Transferring files between a computer and devices	0.819	0.870	0.851	0.929	0.116
<i>Online indicators: Facebook GGI</i>					
FB GG age 18+	0.598	0.840	0.770	0.993	0.231
FB GG age 20–64	0.571	0.866	0.776	1.000	0.240
FB GG age 15–19	0.746	0.992	0.850	1.000	0.209
FB GG age 18–23	0.688	0.910	0.812	1.000	0.214
FB GG age 25–29	0.617	0.869	0.782	0.983	0.227
FB GG age 25–49	0.562	0.877	0.772	1.000	0.248
FB GG age 55–59	0.502	0.951	0.770	1.000	0.277
FB GG age 50–54	0.486	0.926	0.751	1.000	0.289
FB GG age 50+	0.506	0.810	0.730	1.000	0.262
FB GG age 60–64	0.506	0.924	0.768	1.000	0.274
<i>Online indicators: AdWords GGI</i>					
ADW GG age 18+	0.500	0.741	0.709	0.926	0.232
ADW GG age 25+	0.431	0.710	0.670	0.885	0.240
ADW GG age 18–24	0.678	0.887	0.807	1.000	0.220
<i>Offline predictors</i>					
HDI	0.579	0.735	0.703	0.827	0.154
GDP capita PPP 2016 (in \$)	4,072.3	12,220.4	19,558.8	26,686.0	21,548.2
Unemployment gender ratio	0.980	1.200	1.427	1.480	1.243
Mean years of schooling (HDI)	6.10	8.70	8.46	11.20	3.08
Senior managerial work GG	0.308	0.474	0.465	0.610	0.225
Income HDI	0.545	0.707	0.692	0.831	0.181
Secondary education rate HDI	0.370	0.629	0.618	0.887	0.284

**Table 2: Data availability, features, and correlations of the different variables with the ITU Internet Gender Gap Index**

Variable	Number of countries in dataset	Pearson's correlation with ITU Internet GGI	Year
ITU Internet GGI	83	1	Varies (2013–17)
Internet penetration	188	0.695	Annual (2016–17)
Log (GDP per capita)	173	0.680	Annual (2016)
HDI	179	0.737	Annual (2015)
Mean years of schooling (HDI)	179	0.630	Annual (2015)
Educational attainment GG	139	0.630	Annual (2016)
FB GG age 18+	176	0.731	Real time (2018)
FB GG age 20–64	176	0.734	Real time (2018)
FB GG age 25+	176	0.721	Real time (2018)
FB GG age 25–29	177	0.726	Real time (2018)
ADW GG age 18+	166	0.618	Real time (2018)
ADW GG age 25+	172	0.608	Real time (2018)

## 5. Results

### 5.1 Correlation analysis

Table 2 presents different indicators that are the most strongly correlated with the internet GGI. Among offline indicators, the internet penetration rate, log GDP per capita, HDI, and mean years of schooling are positively correlated with the internet GGI. The AdWords and the Facebook GGI variables for different age groups are also strongly correlated with the internet GGI. The ADW GGI has a correlation value with the internet GGI of 0.618 for ages 18+ and 0.608 for ages 25+. The FB GGI measures show a stronger correlation with the internet GGI measures than the ADW GGI ones, with the FB GG for age 18+ and ages 20–64 showing the strongest correlations.

Table 3 presents correlations of the FB and ADW GGI with digital skills gender gap measures from the ITU (DS GGI). The strongest skill-specific measure that is correlated with the FB and ADW indicators is the internet-specific skill of sending emails with attached files, although this is available for only 16 countries, followed by skills linked to using copy/paste tools. These results suggest that Facebook and AdWords gender gap indices could also be used as proxies for gender gaps in low-level digital skills. For high-level skills such as programming the correlation with FB and ADW GGI measures is much weaker, suggesting that these indicators are less effective as proxies for gender gaps in high-level digital skills.

**Table 3: Correlation of the different Digital Skills Gender Gap Indices (DS GGI) with Facebook and AdWords Gender Gap Index**

	Number of countries in dataset	FB GG age 18+	Number of countries in dataset	ADW GG age 18+
DS GGI – Copying or moving file or folder	46	0.731	44	0.604
DS GGI – Using copy and paste tools	32	0.803	31	0.700
DS GGI – Sending emails with attached files	16	0.841	15	0.761
DS GGI – Using basic arithmetic formula in spreadsheet	41	0.550	39	0.479
DS GGI – Connecting & installing new devices	19	0.179	17	0.377
DS GGI – Finding, downloading, and installing software	39	0.452	37	0.393
DS GGI – Creating electronic presentations	46	0.763	43	0.665
DS GGI – Transferring files between computer and a device	47	0.767	44	0.602
DS GGI – Writing computer program using prog. lang.	46	-0.248	43	0.013

## 5.2 Models predicting internet use gender gap

In this section we present results from three types of regression model, the offline-model, the online-model, and the online–offline model (Tables 4 and 5).<sup>15</sup> We also report results comparing the predictive performance of the online and the offline–online models using Facebook and AdWords’ estimates for different age groups. All predictor variables were standardized before fitting the model to make the coefficients of the regression models more comparable with each other. The regression tables report the values of coefficients for the model variables and measures of predictive performance. Since the focus of this analysis is more on the predictive fit of the models and due to collinearity between some of the variables that could influence standard errors and p-values, these measures are not reported in the tables.

### 5.2.1 Offline model

We report a summary of the predictive performance of the offline model in Table 4. The offline model serves as a benchmark to compare the performance of the online and online–offline models, as it shows how well we would be able to predict internet use gender gaps if the online Facebook or AdWords indicators were not available. Various

<sup>15</sup> In Appendix Tables A-1 and A-2 we present results from the same models in which all gender gap indicators are uncapped and log-transformed. The results are substantively similar to those presented in the main text.

offline indicators were selected from the range of available offline indicators by the greedy step-wise forward approach in order to give an Adjusted R-squared value of 0.58. These variables are the country's GDP per capita, Human Development Index (HDI), and unemployment gender ratio, which is the female-to-male ratio of total unemployment. Countries that have higher HDI also have higher gender equality in internet use, while countries that have lower unemployment ratios tend to have higher gender equality in internet use. It is interesting to note that even though we include economic and cultural variables as offline indicators, the economic variables are the ones that are picked up in both the offline and online-offline models (reported later). This suggests that economic factors linked to overall country-level development (e.g., GDP per capita and HDI) appear to be more important for predicting internet use gender gaps than cultural variables such as Hofstede's dimensions, which is consistent with the findings in Garcia et al. (2018). We acknowledge that data on Hofstede's dimensions were available for considerably fewer countries ( $n = 99$ ) than the other offline variables, and this sparsity may also be a factor underlying why it is not predictive of digital gender inequality.

**Table 4: Summary of results for three regression models predicting ITU Internet Gender Gap Index using (1) a single online variable; (2) offline variables and a single online variable; (3) offline variables. Table shows coefficients for standardized values of the explanatory variables. The last row of the table indicates the number of countries in the data for which the model can be used to make predictions.**

	Online model			Online-Offline model			Offline model
	FB	ADW	FB &ADW	FB & offline var	ADW & offline var	FB & ADW & offline var	Offline indicators
Intercept	0.934	0.933	0.933	0.934	0.932	0.932	0.932
GDP capita PPP 2016					-0.029	-0.024	-0.044
HDI				0.049	0.115	0.101	0.110
Mean year schooling HDI					-0.037	-0.037	
Unemployment gender ratio							0.021
FB GG (age 18+)	0.078		0.066	0.047		0.027	
ADW GG (age18+)		0.067	0.016		0.041	0.025	
Adjusted R-squared	0.528	0.374	0.521	0.646	0.684	0.695	0.585
Mean Abs. Error	0.045	0.053	0.045	0.041	0.038	0.038	0.047
SMAPE	5.7%	6.6%	5.8%	5.3%	5.2%	5.3%	6.0%
F-statistics	92.7	46.4	42.3	76.0	41.0	34.8	38.6
Df	81	75	74	80	70	69	77
N	83	77	77	83	75	75	81
Num. pred. countries	176	166	166	174	157	157	162

### 5.2.2 Online model

We first estimated three iterations of the online model using the FB GGI for age group 18+, the ADW GGI for age group 18+, and then both the FB GGI for age group 18+ and the ADW GGI for age group 18+ as predictors of the internet GGI. The results of the online models' predictive performance are shown in Table 4. The positive coefficients on the Facebook and AdWords' GGI variables indicate that gender gaps in Facebook's number of users or Adwords' number of impressions in different countries also capture the gender gap in internet use in those countries. Another way to interpret this is that when women are missing in the Facebook audience estimates and Google impression estimates this serves as a good proxy for the fact that they are not online at all in these countries. The single-variable online model (with intercept) that uses the FB GGI for age group 18+ only as a predictor has marginally the highest adjusted R-squared value and lowest SMAPE of the three online models, but its performance is generally comparable to the FB + ADW model. This is also confirmed with a likelihood ratio test shown in Table A-3 in the Appendix that compares the models on a common subset of countries for which data are available for both models. While the likelihood ratio test allows us to reject the null hypothesis in favor of the alternative hypothesis that the online model with FB and ADW GGI 18+ indicators fits better than the model with ADW alone ( $p < 0.001$ ), the null hypothesis favoring the FB and ADW over FB alone cannot be rejected. Compared with the offline model, the online model that uses the FB GGI for age group 18+ or the FB + ADW 18+ GGI indicators generates more accurate out-of-sample estimates with lower SMAPE.

We further investigated if Facebook and AdWords' indicators for different age groups, either separately or together, could be used to improve the performance of the online model. The findings are presented in Table 5. Several Facebook and AdWords' GGIs for different age groups are positively associated with internet use GGI, and models with multiple age groups tend to show better predictive performance than the single-variable 18+ models in Table 4. AdWords indicators are available for fewer countries for multiple age groups than the Facebook indicators. Even when using multiple age groups, the model performance of the FB and ADW models is comparable to FB only when fitted on a common subset of countries, as shown in Table A-3, although country coverage is poorer with the FB + ADW models compared with the FB model due to AdWords' worse country coverage in different age groups.

**Table 5: Summary of results for three regression models predicting ITU Internet Gender Gap Index using (1) multiple age groups of online variables; (2) offline variables and multiple age groups of online variables; (3) offline variables. All reported coefficients are with standardized values of the predictor variables. The last row of the table indicates the number of countries in the data for which the model can be used to make predictions.**

	Online model			Online–Offline model			Offline model
	FB	ADW	FB & ADW	FB & offline var	ADW & offline var	FB & ADW & offline var	Offline indicators
Intercept	0.934	0.931	0.931	0.934	0.931	0.931	0.932
HDI				0.038	0.061	0.051	0.110
GDP capita PPP 2016							–0.044
Unemployment gender ratio							0.021
FB GG (age 18+)			0.065				
FB GG (age 20–64)				0.151			
FB GG (age 25–29)				–0.038			
FB GG (age 25–49)	0.082						
FB GG (age 55–59)	0.120						
FB GG (age 50–54)			0.099				
FB GG (age 50+)						0.110	
FB GG (age 60–64)	–0.134		–0.110	–0.073		–0.108	
ADW GG (age 18–24)		0.077	0.024		0.047	0.045	
Adjusted R-squared	0.677	0.473	0.717	0.724	0.696	0.796	0.585
Mean Abs. Error	0.042	0.051	0.038	0.038	0.036	0.032	0.047
SMAPE	5.3%	6.3%	5.1%	5.0%	4.7%	4.3%	6.0%
F-statistics	58.4	67.3	47.9	54.8	85.8	73.1	38.6
Df	79	73	70	78	72	70	77
N	83	75	75	83	75	75	81
Num. pred. countries	176	153	153	174	152	152	162

### 5.2.3 Online–Offline model

We examined if the FB GGI for ages 18+ and ADW GGI for ages 18+ could be combined with offline predictors either separately or jointly to improve predictive performance. The results from these online–offline models are reported in Table 4. Offline variables selected by the step-wise approach include variables linked to economic development such as GDP per capita, HDI, and the HDI sub-index for mean years of schooling. The cultural variables are not selected in any of the online–offline models in the same way as the offline models. The online–offline models show better predictive performance than the online models and the offline model. Within the online–offline models in Table 4 the online–offline model with Facebook and AdWords indicators combined performs the best. This conclusion is also supported by the likelihood ratio test fitted on a common subset of countries for which data are available to fit both models reported in the Appendix in Table A-4. Country coverage, however,

is again better for the online–offline model with FB indicators than the models that rely on AdWords indicators.

We further examined if additional Facebook and AdWords indicators for different age groups could be used to improve the performance of online–offline models beyond the global (18+) FB GGI and ADW GGI variables. These findings are displayed in Table 5. On the whole, the online–offline model that uses both the FB GGI and the ADW GGI for different age groups combined with offline indicators performs the best of all the online, online–offline, and offline models reported in both Tables 4 and 5. This model produces the highest adjusted R-squared and the lowest mean abs. error and SMAPE, and improves on the online–offline model reported in Table 4 that uses the FB and ADW GGI for ages 18+ with offline indicators. This is supported by a Vuong test comparing the two (non-nested) FB+ADW+offline models from Tables 4 and 5, which is reported in the Appendix in Table A-4 where both models are fitted on a common subset of countries. However, this model is able to predict internet access gender gaps for fewer countries than the online–offline model with FB indicators in Table 5. A further comparison of the online–offline FB model using multiple age groups in Table 5 with the online–offline FB + ADW model in Table 5 with a Vuong test does not provide support for favoring the online–offline FB + ADW model over the online–offline with FB only model in Table 5.

### **5.3 Predicting gender gaps in digital skills**

As the correlations in Table 3 indicate, the usefulness of the gender gap measures from Facebook and Google advertising extends beyond predicting gender gaps in internet use; they can also provide useful proxies for estimating gender gaps in a variety of digital skills, which are another important dimension of digital gender gaps. Unlike self-reported indicators in the ITU survey data, these indicators are a different kind of measure of digital skills as they capture use of specific platforms such as Facebook and Google, and thus indirectly capture users' digital skills that are necessary to make use of these services.

Table 6 provides a summary of regression models providing the best performance that were fitted to predict a variety of different digital skill gender gaps using a combination of the online FB and ADW GGI and offline variables. The digital skills shown in the table range from basic editing skills, such as using copy and paste tools and file transfer skills, to relatively more advanced skills, such as numeric skills involving the use of a spreadsheet and computer programming.



**Table 6: Summary of results for regression models predicting different ICT skills gender gaps: (1) Using copy and paste tools, (2) Using basic arithmetic formula in a spreadsheet, (3) Writing computer programs in programming language, (4) Transferring files between a computer and devices. All reported coefficients are with standardized values of the predictor variables. The last row of the table indicates the number of countries in the data for which the model can be used to make predictions.**

	(1) Copy & paste tools	(2) Formula in spreadsheet	(3) Programming language	(4) File transfer
Intercept	0.903	0.860	0.472	0.851
FB GG (age 15–19)		0.069		0.042
FB GG (age 18–23)	0.100			
FB GG (age 50–54)				0.058
ADW GG (age 25+)			0.091	
Senior managerial work GG		0.038		
Income HDI	0.034			
Secondary Educ. rate HDI			–0.150	
Adjusted R-squared	0.727	0.505	0.325	0.671
Mean Abs. Error	0.056	0.073	0.122	0.054
SMAPE	7.81%	9.33%	28.07%	7.41%
F-statistics	42.31	21.38	11.85	47.85
Df	29	38	43	44
N	32	41	46	47
Num. pred. countries	174	118	154	177

For three of the four skill measures, FB GGI measures are selected as the online indicators, the exception being writing a computer program, for which the ADW measure is selected. The offline indicators selected include development measures such as the HDI pertaining to income and education as well as a variable pertaining to gender gaps in holding senior positions in the workplace (Senior managerial GG). As in the internet GGI models the cultural variables are not selected, indicating again the importance of economic factors in predicting digital gender gaps and improving the predictive power of online variables. For three of the four skill outcomes, online indicators combined with offline indicators provide the models with the best predictive fit. For skills related to transferring files between a computer and a device, the step-wise selection algorithm only picks up online indicators.

Some digital skills are more strongly predicted by the variables in the data than others. This can be seen in Table 6, where the models have relatively higher adjusted R-squared and smaller errors for low-level skills such as using copy and paste tools (adj. R-squared 0.727) and file transfer skills (adj. R-squared 0.671) than for high-level skills such as programming (adj. R-squared 0.325) or using a basic arithmetic formula in a spreadsheet (adjusted R-squared 0.505). The coefficients of the predictor variables are positive, which indicates that when we observe a smaller gender disparity on Facebook

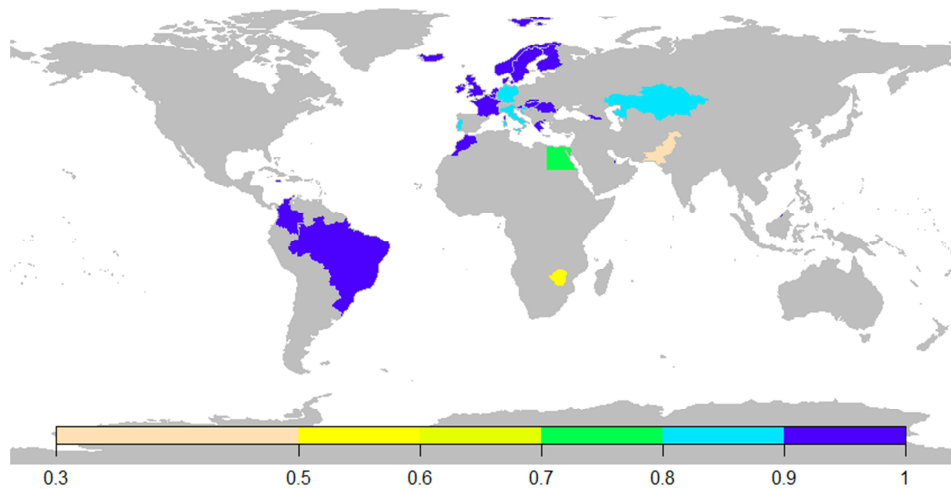
or on Google AdWords we also expect to observe a smaller gender disparity across different types of skills.

A plausible explanation of the differences in the ability of the FB and ADW indicators to serve as proxies for low-level rather than high-level skills is the type of skill required to use these platforms. For example, someone using Facebook or Google may need to be able to perform tasks such as copying and pasting a search term on Google, or moving files around when uploading an image on Facebook. As a result, we see stronger correlations between these types of skill and the gender gap on Facebook than we see for more high-level skills such as programming. We find that FB GGI indicators perform better than ADW GGI indicators for predicting both internet use and low-level skills. This could arise from the fact that FB indicators provide counts of users rather than impressions, which is what is provided by AdWords. Alternatively, being a Facebook user might be a good proxy for basic internet use and digital skills, but as ADW captures a range of different platforms, more active use of these platforms, for example, by male users, could make this a less effective proxy for general internet use. As the ITU measure that we are predicting captures general internet use rather than frequency, intensity, or specific type of use, the Facebook indicators appear to provide better proxies for this than the AdWords indicators.

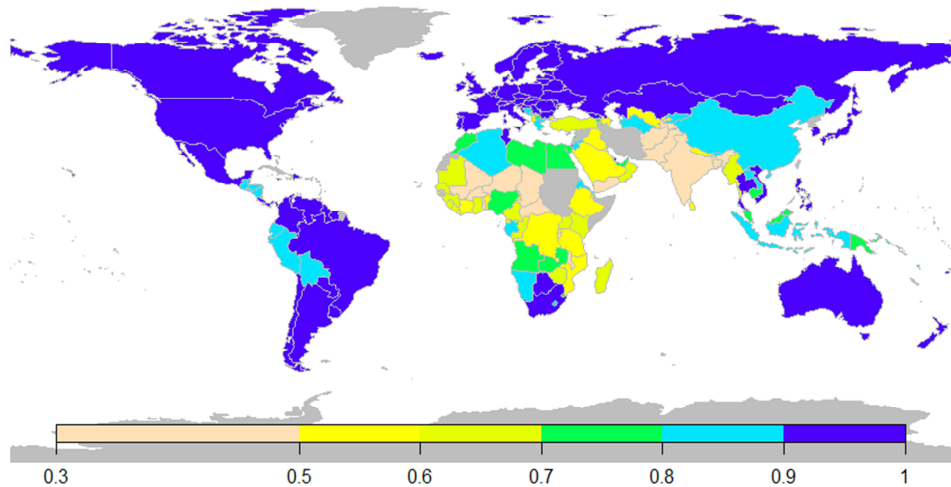
Figure 3a shows the gender gap measure for using copy and paste tools from the ITU data and Figure 3b shows its value as predicted by the model in Table 6. As can be seen, for countries where data are available, the model predictions largely agree with the data. However, the coverage of countries is greatly enhanced when using the online data, especially for Africa where ITU data are sparse. Figure 4 shows a scatter plot of the gender gap for using copy and paste tools from ITU data against the predicted values from the model. Although countries generally align along the  $x = y$  diagonal, with no systematic skew in under- or over-prediction, there are some outliers. The two largest residuals in terms of over-prediction where the model predicts more gender equality than ITU indicators are Pakistan and Croatia. Conversely, Morocco and Brunei are the two largest residuals in terms of under-prediction, i.e., where the model predicts less gender equality than indicated by the ITU data.

**Figure 3: Maps of the digital skill (using copy and paste tools) gender gap index (DG GGI) computed using ITU data (panel a) and as predicted by the model (panel b)**

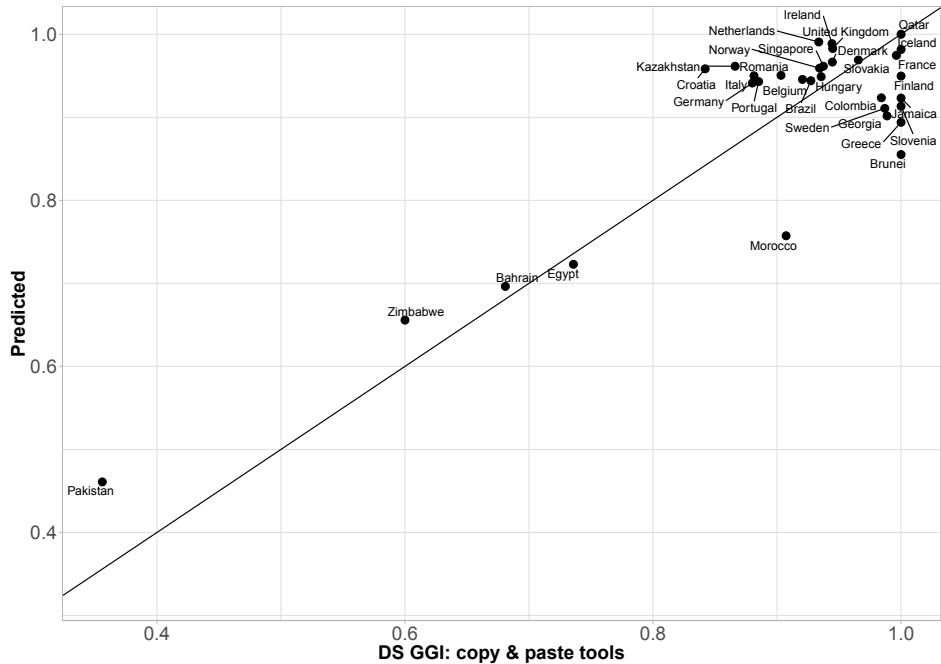
(a) DS GGI – Using copy and paste tools computed using data from International Telecommunications Union (ITU)



(b) Using copy and paste tools predicted using the online-offline model (Facebook data combined with offline indicators)



**Figure 4:** Scatter plot showing DS GGI – Using copy and paste tools computed using ITU data (x-axis) against the values predicted from the model using Facebook data combined with offline indicators (y-axis). The plotted line is the diagonal  $x=y$  line.



## 6. Discussion

Women’s equal participation in the digital society is considered integral to achieving global goals related to gender equality. The ITU has highlighted the importance of “putting in place data, monitoring and evaluation tools around gender equality and ICT, including for measurement of access and use” to realize these goals (International Telecommunication Union 2015). This study explores how anonymous, aggregate data from the online populations of Google and Facebook can help with this endeavor.

Our study expands in several ways on prior research in Fatehkia, Kashyap, and Weber (2018) that uses Facebook’s advertisement audience estimates to predict gender gaps in internet use. First, we evaluated the potential of another novel data source – Google’s advertisement impression estimates (AdWords) – to predict gender gaps in

internet use around the world. To the best of our knowledge, this paper is the first to use impressions estimates from Google AdWords to monitor any target related to the Sustainable Development Goals (SDGs). Second, we explored whether Facebook and AdWords' measures for different age groups could be combined to improve models for predicting gender gaps in internet use. Third, we use the latest ground truth ITU data to compute gender gaps in specific types of digital skills and to examine their relationship with Facebook and AdWords' derived measures.

The prediction results are very promising, as the online model using Facebook and AdWords' gender gap measures for different age groups is able to explain 72% of the variance in the ground truth of the ITU Internet Gender Gap Index, showing a slight improvement over the online model reported in Fatehikia, Kashyap, and Weber (2018), which explained 69% of the variance.<sup>16</sup> By comparison, the offline model only explains 59% of the variance. Furthermore, the online–offline model that uses both the Facebook and AdWords' gender gap measures for different age groups combined with offline indicators has the best model fit, explaining around 80% of the variance in the ground truth. This supports our approach of integrating both Facebook and AdWords' gender gap measures to the regression model, as using only offline development indicators might not be enough to provide good predictions of internet-use gender gaps. In addition, our approach shows that the predictive performance is improved by combining Facebook and AdWords' gender gap measures for different age groups. Nevertheless, we note that the coverage of AdWords for different age groups is worse than that of Facebook, so drawing on AdWords data we are able to make predictions for fewer countries than when relying on Facebook alone. In terms of robustness and generalizability, however, drawing on both Google and Facebook data gives us different ways of operationalizing our indicators of interest. Thus, if the user behavior on one of the platforms changes radically the combined prediction will hopefully remain somewhat stable and guard against outliers.

Going beyond modeling mere binary internet use-or-not, our results also demonstrate a strong relationship between gender gaps in digital skills and gender gaps in Facebook and Adwords' estimates. Specifically, we found that gender gaps in digital skills such as sending e-mail with attached files or using copy and paste tools are associated with a gender gap in Facebook and Adwords' estimates. In general, we found that Facebook variables are better able to predict low-level skills, which could be linked to both the type of digital use that having a Facebook account is proxy for and the quantity provided by the Facebook marketing platform compared with AdWords (users versus impressions). Further research could use Facebook and AdWords' data to

---

<sup>16</sup> Note that the online model with Facebook GGI 18+ performs worse with the latest round of ITU (2018) used in this study than with the ITU (2016) data used in Fatehikia, Kashyap, and Weber (2018).

measure gender gaps in digital skills by filtering advertising-reach estimates based on education or specific audience interests.

Using aggregate, anonymous online-advertising audience estimates to predict digital gender gaps has several advantages over traditional approaches. First, data from online advertising platforms can be collected regularly, enabling predictions to be made on, say, a monthly basis. Second, large internet platforms such as the Facebook and Google have a global reach and their data can be collected for a large number of countries, enabling predictions of internet use and digital skill gender gaps for most countries in the world. Here, the biggest gain in coverage is for low- and lower-middle-income countries where online data enable us to generate estimates for 64 countries, compared with 16 in the ITU data. Third, the gender gap in Facebook and AdWords can be disaggregated by different sociodemographic characteristics such as age, gender, language, and location, which is particularly useful for large countries like India. Leveraging these additional characteristics provides further avenues for extending this work. In this paper we have used Facebook and Google gender gap indicators as predictor variables to examine their validity against survey-based measures of digital gender inequality. Further extensions of this work could analyze these online gender gaps as outcomes to understand what factors explain their variation, and these data could be routinely collected to examine how they vary over time. These online platforms could themselves be used to field surveys and estimate digital inequalities, using approaches such as that in Feehan and Cobb (2019).

Nevertheless, our approach of combining Facebook and AdWords' gender gap measures to predict gender gaps in internet use has several limitations. First, Facebook and Google do not provide documentation that explains how their algorithms estimate the number of users or the number of impressions. Hence, these estimates are sensitive to changes in the design of the black-box algorithms. Second, no online advertising platform that we are aware of provides historic estimates, disaggregated by period, for the reach of a to-be-launched ad campaign. This makes it difficult to evaluate changes in the model fit over time. Indeed, we found that the Facebook and AdWords estimates have a higher correlation with the 2016 ITU data than with the 2018 ITU data. Also, the 2018 ITU data are based on gender-disaggregated data on internet use from 2013–2017, depending on the country, whereas the Facebook and the AdWords data are more recent. As Facebook and Google do not provide historic data, we need to continue collecting data prospectively to be able to track changes over time and to update our regression models. Nevertheless, our approach of combining Facebook and AdWords' gender gap measures to predict digital gender gaps proves that these data sources can help with up-to-date monitoring of SDG targets linked to digital gender equality. Finally, in extending our approach to monitoring other SDGs and other aspects of human development, it has to kept in mind that not everyone is on the internet, let alone

on Facebook or Google. One contribution of our work here has been to highlight how women are significantly underrepresented on two of the largest online platforms, Google and Facebook, particularly in several countries in sub-Saharan Africa and South Asia. Whereas in this particular study the main signal used was the absence of users, in other studies it would be important to be cognisant of this bias when using these online data sources to measure social indicators, or to recruit participants for research studies. The underrepresentation of women on Facebook and Google, and more broadly online, is also significant for development practitioners and NGOs who are deploying media and information campaigns, as well as other social interventions, on social media and other online platforms. Moving forward, this work provides an example of how online data can play a complementary role in advancing the ‘data revolution’ for sustainable development (IEAG 2014). Online data sources have biases that are important to consider, but by using them in conjunction with traditional data from surveys and censuses, population scientists can contribute to the task of evaluating and augmenting their potential.

## **7. Acknowledgements**

The research for this paper was conducted as a part of the project ‘The Digital Traces for the Gender Digital Divide’ that received funding from Data2X, an initiative of the United Nations Foundation (Grant No. UNF-17-936, PI: R. Kashyap). We are grateful to the Data2X team for their support throughout the project. We thank Ian Knowles for technical support. Earlier versions of this paper were presented at the Population Association of America 2019 annual meeting and the International Conference for Computational Social Science 2019. We thank participants at both conferences for helpful feedback. Ridhi Kashyap acknowledges support from the Leverhulme Trust within the Leverhulme Centre for Demographic Science.

## References

- Abu-Shanab, E. and Al-Jamal, N. (2015). Exploring the gender digital divide in Jordan. *Gender, Technology and Development* 19(1): 91–113. doi:10.1177/0971852414563201.
- Aker, J.C. and Mbiti, I.M. (2010). Mobile phones and economic development in Africa. *Journal of Economic Perspectives* 24(3): 207–232. doi:10.1257/jep.24.3.207.
- Antonio, A. and Tuffley, D. (2014). The gender digital divide in developing countries. *Future Internet* 6: 673–687. doi:10.3390/fi6040673.
- Araújo, M., Mejova, Y., Weber, I., and Benevenuto, F. (2017). Using Facebook ads audiences for global lifestyle disease surveillance: Promises and limitations. In: *WebSci '17: Proceedings of the 2017 ACM on Web Science Conference*. New York: Association for Computing Machinery: 253–257. doi:10.1145/3091478.3091513.
- Beck, F., Richard, J.-B., Nguyen-Thanh, V., Montagni, I., Parizot, I., and Renahy, E. (2014). Use of the internet as a health information resource among French young adults: Results from a nationally representative survey. *Journal of Medical Internet Research* 16(5): e128. doi:10.2196/jmir.2934.
- Benita, F.C. (2019). A new measure of transport disadvantage for the developing world using free smartphone data. *Social Indicators Research* 145: 415–435. doi:10.1007/s11205-019-02107-x.
- Billari, F.C., Giuntella, O., and Stella, L. (2019). Does broadband internet affect fertility? *Population Studies* 73(3): 297–316. doi:10.1080/00324728.2019.1584327.
- Billari, F.C., Rotondi, V., and Trinitapoli, J. (2020). Mobile phones, digital inequality, and fertility: Longitudinal evidence from Malawi. *Demographic Research* 42(37): 1057–1096. doi:10.4054/DemRes.2020.42.37.
- Blumenstock, J., Cadamuro, G., and On, R. (2015). Predicting poverty and wealth from mobile phone metadata. *Science* 350(6264): 1073–1076. doi:10.1126/science.aac4420.
- Broadband Commission (2013). Doubling digital opportunities: Enhancing the inclusion of women and girls in the information society. Technical report, UNESCO, ITU. <https://www.broadbandcommission.org/Documents/publication/s/bb-doubling-digital-2013.pdf>.



- Choi, H. and Varian, H. (2012). Predicting the present with Google trends. *Economic Record* 88: 2–9. doi:10.1111/j.1475-4932.2012.00809.x.
- Chunara, R., Bouton, L., Ayers, J.W., and Brownstein, J.S. (2013). Assessing the online social environment for surveillance of obesity prevalence. *PLoS One* 8(4): 1–8. doi:10.1371/journal.pone.0061373.
- Detting, L.J. (2017). Broadband in the labor market: The impact of residential high-speed internet on married women’s labor force participation. *ILR Review* 70(2): 451–482. doi:10.1177/0019793916644721.
- di Bella, E., Leporatti, L., and Maggino, F. (2016). Big data and social indicators: Actual trends and new perspectives. *Social Indicators Research* 135: 869–878. doi:10.1007/s11205-016-1495-y.
- DiMaggio, P., Hargittai, E., Celeste, C., and Shafer, S. (2004). From unequal access to differentiated use: A literature review and agenda for research on digital inequality. Technical report, Russell Sage Foundation. <https://www.russellsage.org/sites/all/files/u4/DiMaggio%20et%20al.pdf>.
- Elvidge, C.D., Sutton, P.C., Ghosh, T., Tuttle, B.T., Baugh, K.E., Bhaduri, B., and Bright, E. (2009). A global poverty map derived from satellite data. *Computers and Geosciences* 35(8): 1652–1660. doi:10.1016/j.cageo.2009.01.009.
- European Parliament (2018). The underlying causes of the digital gender and possible solutions for enhanced digital inclusion of women and girls. Policy PE 604.940, Policy Department for Citizen’s Rights and Constitutional Affairs. [http://www.europarl.europa.eu/RegData/etudes/STUD/2018/604940/IPOL\\_STU\(2018\)604940\\_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/STUD/2018/604940/IPOL_STU(2018)604940_EN.pdf).
- Fatehkia, M., Kashyap, R., and Weber, I. (2018). Using Facebook ad data to track the global digital gender gap. *World Development* 107: 189–209. doi:10.1016/j.worlddev.2018.03.007.
- Feehan, D.M. and Cobb, C. (2019). Using an online sample to estimate the size of an offline population. *Demography* 56(6): 2377–2392. doi:10.1007/s13524-019-00840-z.
- Garcia, D., Kassa, Y.M., Cuevas, A., Cebrian, M., Moro, E., Rahwan, I., and Cuevas, R. (2018). Analyzing gender inequality through large-scale Facebook advertising data. *Proceedings of the National Academy of Sciences* 115(27): 6958–6963. doi:10.1073/pnas.1717781115.

- Gurumurthy, A. and Chami, N. (2014). Gender equality in the information society. Technical report, IT for Change. [https://assets.publishing.service.gov.uk/media/57a089b0e5274a31e0000202/61047\\_GOKH\\_Gender-and-ICTs-briefing-2014.pdf](https://assets.publishing.service.gov.uk/media/57a089b0e5274a31e0000202/61047_GOKH_Gender-and-ICTs-briefing-2014.pdf).
- Hafkin, N.J. and Huyer, S. (2007). Women and gender in ICT statistics and indicators for development. *Information Technologies and International Development* 4(2): 25–41. doi:10.1162/itid.2008.00006.
- Hargittai, E. (2002). Second-level digital divide: Differences in people's online skills. *First Monday* 7(4). doi:10.5210/fm.v7i4.942.
- Hargittai, E. and Shafer, S. (2006). Differences in actual and perceived online skills: The role of gender. *Social Science Quarterly* 87(2): 432–448. doi:10.1111/j.1540-6237.2006.00389.x.
- Hilbert, M. (2011). Digital gender divide or technologically empowered women in developing countries? A typical case of lies, damned lies, and statistics. *Women's Studies International Forum* 34(6): 479–489. doi:10.1016/j.wsif.2011.07.001.
- Hjort, J. and Poulsen, J. (2019). The arrival of fast internet and employment in Africa. *American Economic Review* 109(3): 1032–1079. doi:10.1257/aer.20161385.
- Hofstede, G. (2011). Dimensionalizing cultures: The Hofstede model in context. *Online Readings in Psychology and Culture* 2(1): 8. doi:10.9707/2307-0919.1014.
- IEAG (2014). A world that counts: Mobilising the data revolution for sustainable development. Technical report <https://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf>.
- Intel (2012). Women and the Web. Technical report. <https://www.intel.com/content/dam/www/public/us/en/documents/pdf/women-and-the-web.pdf>.
- International Telecommunication Union (2015). Action plan to close the digital gender gap. Policy, International Telecommunication Union. <https://www.itu.int/en/action/gender-equality/Documents/ActionPlan.pdf>.
- International Telecommunication Union (2016). Core list of ICT indicators. Retrieved from the International Telecommunication Union website, [https://www.itu.int/en/ITU-D/Statistics/Documents/coreindicators/Core-List-of-Indicators\\_March2016.pdf](https://www.itu.int/en/ITU-D/Statistics/Documents/coreindicators/Core-List-of-Indicators_March2016.pdf).

- International Telecommunication Union (2017). Fast-forward progress leveraging tech to achieve the global goals. [https://www.itu.int/en/sustainable-world/Documents/Fast-forward\\_progress\\_report\\_414709%20FINAL.pdf](https://www.itu.int/en/sustainable-world/Documents/Fast-forward_progress_report_414709%20FINAL.pdf).
- International Telecommunication Union (2018). World telecommunication/ICT indicators database 2018. Data retrieved from the International Telecommunication Union website. <https://www.itu.int/en/ITU-D/Statistics/Pages/publications/wtid.aspx>.
- IUSSP (2015). The IUSSP on a data revolution for development. *Population and Development Review* 41(1): 172–177. doi:10.1111/j.1728-4457.2015.00041.x.
- Lagan, B.M., Sinclair, M., and Kernohan, W.G. (2011). What is the impact of the internet on decision-making in pregnancy? A global study. *Birth* 38(4): 336–345. doi:10.1111/j.1523-536X.2011.00488.x.
- Letouze, E. and Jutting, J. (2014). Official statistics, big data and human development: Towards a new conceptual and operational approach. Technical report, Data Pop Alliance and PARIS21. <https://www.odi.org/sites/odi.org.uk/files/odi-assets/events-documents/5161.pdf>.
- Lund, S., Nielsen, B.B., Hemed, M., Boas, I.M., Said, A., Said, K., Makungu, M.H., and Rasch, V. (2014a). Mobile phones improve antenatal care attendance in Zanzibar: A cluster randomized controlled trial. *BMC Pregnancy and Childbirth* 14(1): 29. doi:10.1186/1471-2393-14-29.
- Lund, S., Rasch, V., Hemed, M., Boas, I.M., Said, A., Said, K., Makundu, M.H., and Nielsen, B.B. (2014b). Mobile phone intervention reduces perinatal mortality in Zanzibar: Secondary outcomes of a cluster randomized controlled trial. *JMIR Health and Health* 2(1): e15. doi:10.2196/mhealth.2941.
- Mao, H., Shuai, X., Ahn, Y.-Y., and Bollen, J. (2015). Quantifying socio-economic indicators in developing countries from mobile phone communication data: Applications to Côte d’Ivoire. *EPJ Data Science* 4(15). doi:10.1140/epjds/s13688-015-0053-1.
- Mejova, Y., Gandhi, H.R., Rafaliya, T.J., Sitapara, M.R., Kashyap, R., and Weber, I. (2018). Measuring subnational digital gender inequality in India through gender gaps in Facebook use. In: *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*. New York: Association for Computing Machinery. doi:10.1145/3209811.3212698.
- Nuti, S.V., Wayda, B., Ranasinghe, I., Wang, S., Dreyer, R.P., Chen, S.I., and Murugiah, K. (2014). The use of Google trends in health care research: A

- systematic review. *PloS One* 9(10): e109583. doi:10.1371/journal.pone.0109583.
- Ono, H. and Zavodny, M. (2007). Digital inequality: A five country comparison using micro-data. *Social Science Research* 36(3): 1135–1155. doi:10.1016/j.ssr.2006.09.001.
- Rampazzo, F., Zagheni, E., Weber, I., Testa, M.R., and Billari, F.C. (2018). *Mater certa est, pater numquam: What can Facebook advertising data tell us about male fertility rates?* International Conference on Web and Social Media (ICWSM).
- Resce, G. and Maynard, D. (2018). What matters most to people around the world? Retrieving better life index priorities on Twitter. *Technological Forecasting and Social Change* 137: 61–75. doi:10.1016/j.techfore.2018.06.044.
- Robinson, L., Cotten, S.R., Ono, H., Quan-Haase, A., Mesch, G., Chen, W., Schulz, J., Hale, T.M., and Stern, M.J. (2015). Digital inequalities and why they matter. *Information, Communication and Society* 18(5): 569–582. doi:10.1080/1369118X.2015.1012532.
- Rotondi, V., Kashyap, R., Pesando, L.M., Spinelli, S., and Billari, F.C. (2020). Leveraging mobile phones to attain sustainable development. *Proceedings of the National Academy of Sciences* 117(24): 13413–13420. doi:10.1073/pnas.1909326117.
- Salganik, M.J. (2017). *Bit by bit: social research in the digital age*. Princeton, NJ: Princeton University Press.
- Scheerder, A., van Deursen, A., and van Dijk, J. (2017). Determinants of internet skills, uses and outcomes: A systematic review of the second- and third-level digital divide. *Telematics and Informatics* 34(8): 1607–1624. doi:10.1016/j.tele.2017.07.007.
- Suri, T. and Jack, W. (2016). The long-run poverty and gender impacts of mobile money. *Science* 354(6317): 1288–1292. doi:10.1126/science.aah5309.
- Tan, S., Zhao, Y., and Huang, W. (2019). Neighborhood social disadvantage and bicycling behavior: A big data-spatial approach based on social indicators. *Social Indicators Research* 145: 985–999. doi:10.1007/s11205-019-02120-0.
- United Nations (2016). Global sustainable development report. Technical report, Division of Economic and Social Affairs.
- United Nations Population Division (2017). World population prospects: The 2015 revision. <https://esa.un.org/unpd/wpp/Download/Standard/Population/>.

- Unwin, P. and Unwin, T. (2009). *ICT4D: Information and communication technology for development*. Cambridge: Cambridge University Press.
- Weber, I., Kashyap, R., and Zagheni, E. (2018). Using advertising audience estimates to improve global development statistics. *ITU Journal: ICT Discoveries* 1(2). [https://www.itu.int/dms\\_pub/itu-s/opb/journal/S-JOURNAL-ICTF.VOL1-2018-2-P04-PDF-E.pdf](https://www.itu.int/dms_pub/itu-s/opb/journal/S-JOURNAL-ICTF.VOL1-2018-2-P04-PDF-E.pdf).
- World Economic Forum (2016). Global gender gap report 2016. Technical report, World Economic Forum. <http://wef.ch/1yrt8iq>.
- WWW Foundation (2015). Women's Rights online: Translating access into empowerment. <https://webfoundation.org/research/womens-rights-online-2015/>.
- Zagheni, E., Weber, I., and Gummadi, K. (2017). Leveraging Facebook's advertising platform to monitor stocks of migrants. *Population and Development Review* 43(4): 721–734. doi:10.1111/padr.12102.

## Appendix A: Log transforming the gender gap indicators

Tables A-1 and A-2 present results from the same models as Tables 4 and 5 but with all gender gap indicators uncapped and log-transformed. For comparability with the results reported in Tables 4 and 5, the Mean Abs. Error and the SMAPE were computed after transforming the predicted and ground truth variables back by applying the inverse log function and capping the variables at 1.

**Table A-1: Summary of results for three regression models predicting ITU internet Gender Gap Index using using (1) a single online variable; (2) offline variables and a single online variable; (3) offline variables. Table shows coefficients for standardized values of the explanatory variables.**

	Online model			Online–offline model			Offline model
	FB	ADW	FB & ADW	FB & offline var	ADW & offline var	FB & ADW & offline var	Offline indicators
Intercept	-0.074	-0.076	-0.076	-0.074	-0.077	-0.077	-0.076
GDP capita PPP 2016					-0.048	-0.033	-0.075
HDI				0.056	0.159	0.124	0.163
Mean year schooling HDI					-0.046	-0.049	
Unemployment ratio							0.036
FB GG (age 18+)	0.119		0.114	0.083		0.073	
ADW GG (age18+)		0.098	0.010		0.058	0.016	
Adjusted R-squared	0.599	0.383	0.590	0.672	0.654	0.705	0.578
Mean Abs. Error	0.043	0.052	0.043	0.042	0.042	0.040	0.051
SMAPE	5.34%	6.39%	5.63%	5.27%	5.56%	5.51%	6.51%
F-statistics	123.6	48.2	55.8	85.1	35.9	36.4	37.5
Df	81	75	74	80	70	69	77
N	83	77	77	83	75	75	81

**Table A-2: Summary of results for three regression models predicting ITU internet Gender Gap Index using using (1) multiple age groups of online variables; (2) offline variables and multiple age groups of online variables; (3) offline variables. All reported coefficients are with standardized values of the predictor variables.**

	Online model			Online–offline model			Offline model
	FB	ADW	FB & ADW	FB & offline var	ADW & offline var	FB & ADW & offline var	Offline indicators
Intercept	-0.074	-0.078	-0.078	-0.074	-0.078	-0.078	-0.076
HDI				0.043	0.089	0.059	0.163
GDP capita PPP 2016							-0.075
Unemployment ratio							0.036
FB GG (age 18+)			0.056				
FB GG (age 20–64)				0.213			
FB GG (age 25–29)				-0.047			
FB GG (age 25–49)	0.124						
FB GG (age 55–59)	0.150						
FB GG (age 50–54)			0.185				
FB GG (age 50+)						0.208	
FB GG (age 60–64)	-0.174		-0.158	-0.096		-0.182	
ADW GG (age 18–24)		0.103	0.030		0.062	0.048	
Adjusted R-squared	0.727	0.404	0.783	0.757	0.644	0.802	0.578
Mean Abs. Error	0.045	0.055	0.041	0.041	0.040	0.036	0.051
SMAPE	5.59%	6.84%	5.26%	5.18%	5.18%	4.73%	6.51%
F-statistics	73.8	51.1	67.8	64.8	67.8	76.0	37.5
Df	79	73	70	78	72	70	77
N	83	75	75	83	75	75	81

## Appendix B: Comparing models with likelihood ratio test

Tables A-3 and A-4 compare the performance of the different Online and Online–Offline models in Tables 4 and 5, using the likelihood ratio test or the Vuong test for non-nested models. In order to run the tests, both models being compared were fitted on a common subset of countries for which data were available for the variables of both models; as a result, the table also reports the adj. R2 and number of data points on which the models being compared were fitted.

**Table A-3: Online models: Likelihood-ratio tests/Vuong tested for Non-nested models.**

Model 1	Model 1 adj. $R^2$	Model 2	Model 2 adj. $R^2$	N	Models Nested? (Y/N)	Test statistic	p-value
Table 4: FB	0.518	Table 4: FB+ADW	0.521	77	Y	1.439	0.23
Table 4: ADW	0.374	Table 4: FB+ADW	0.521	77	Y	21.627	3.31e-6
Table 5: FB	0.684	Table 5: FB+ADW	0.717	75	N	-1.513	0.065
Table 5: ADW	0.473	Table 5: FB+ADW	0.717	75	Y	49.912	8.34e-11
Table 4: FB+ADW	0.545	Table 5: FB+ADW	0.719	71	N	-2.029	0.021

**Table A-4: Online–Offline models: Likelihood ratio tests/Vuong tested for non-nested models**

Model 1	Model 1 adj. $R^2$	Model 2	Model 2 adj. $R^2$	N	Models Nested? (Y/N)	Test statistic	p-value
Table 4: Off.+FB	0.662	Table 4: Off.+FB+ADW	0.695	75	Y	10.988	0.012
Table 4: Off.+ADW	0.684	Table 4: Off.+FB+ADW	0.695	75	Y	3.871	0.049
Table 5: Off.+FB	0.740	Table 5: Off.+FB+ADW	0.796	75	N	-0.939	0.174
Table 5: Off.+ADW	0.696	Table 5: Off.+FB+ADW	0.796	75	Y	31.912	1.18e-7
Table 4: Off.+FB+ADW	0.717	Table 5: Off.+FB+ADW	0.801	69	N	-1.829	0.034