Analysis of gender gaps and differences in access to and use of digital technologies in small areas of South Africa: An application of the spatial Fay-Herriot model

YA Shiferaw

Dep. of Statistics, University of Johannesburg

ERSA Workshop 30 September 2022

Outline

1. Digital divides
2. Methodology
   - The FH area-level model
   - The spatial FH model
   - Mean squared error
3. Data Sources
4. Results
   - Testing digital inequality between men and women

Digital divides

Digital divide: what is it?

Digital divides

- The term digital divide has different meanings to different people (Ritzhaupt et al., 2013).
- This term became part of the researchers' vocabulary around 1995s (Wilhelm et al., 2002).
- It is used to describe the uneven distribution of the ability to access and use ICT between individuals due to **socio-economic status, ethnicity and gender** (van Dijk, 2006).
- The impact of digital technologies is mixed, uneven and unrealized (World Bank, 2016).
- This uneven distribution of the gain from digital technologies is rooted in gender inequalities and more research should be conducted to understand it in its specific contexts.

Digital divides ···

Over the last two decades:

- Digitalization has created new consumer, business, and employee behaviours, work patterns, and communication styles (Myovellaa et al., 2020).
- ICT, especially the internet and mobile phone technologies, has increased these patterns, creating new products, processes, market channels, and organizational complexities (Myovellaa et al., 2020).
- Technology developments and ICT changes are considered indispensable to economic prosperity.
- Various economic sectors have been supported by the growth of ICTs, online marketing platforms, and sophisticated digital devices (Zhu, 2021).
- ICT can help developing nations achieve higher growth rates (Acemoglu & Robinson, 2013).

Digital divides · · ·

Why digital gender divides?

- According to the United Nations Conference on Trade and Development (UNCTAD), women are severely under-represented in the digital economy.
- They also receive inadequate funding in this sector. Those facts are from sub-Saharan Africa and East Asia (UNCTAD, 2018).
- In ICT jobs, top management, and academic careers, women are under-represented, according to the Organisation for Economic Co-operation and Development (OECD) report.
- According to the report, men are four times likelier to specialize in ICT than women (OECD., 2018).

Digital divides · · ·

Digital gender divides in South Africa:

- South Africa is no different from most countries regarding the gender digital divide.
- Women have fewer mobile phones, computers, and access to the internet than men.
- Table 1 illustrates the gender gap in internet access in South Africa in 2011.
- This table shows that 72.7% of men and 27.3% of women used internet services at home, while 58.2% of men and 41.8% of women used internet services from their cell phones, 66.8% of men and 33.2% of women using internet services at work.
- The gender gap in Internet access at home was 45.4% in the digital sector, which indicates a significant gender gap.

Digital divides · · ·

Table 1: A summary of the gender gap in access to the internet and cell phone ownership: Population Census 2011. The proportion figures were calculated based on the number of Internet users and cellphone owners in South Africa.

		Men (%)	Women (%)	Diff: Men - Women
Internet use	From home	72.7	27.3	45.4
	From Cellphone	58.2	41.8	16.4
	From work	66.8	33.2	33.6
Own Cellphone		58.4	41.6	16.8

4IR

Indeed, as the Fourth Industrial Revolution is now underway, the analysis of digital statistics can provide policymakers with a clear picture of South Africa's preparation level to welcome this revolution.

Digital divides · · ·

Information at the national and provincial levels in South Africa:

- For example, Stats SA offers a wealth of information at the national and provincial levels in South Africa.
- The main reason is that household surveys, the best available data source on household income and consumption, digital statistics, health outcomes, and so on, are only representative at the provincial level.
- Examples of household surveys:
  - General Household Survey
  - Labour Statistics and Activities Surveys
  - Living Conditions Survey
  - Governance,Public Safety and Justice Survey
  - Victims of Crime Survey
  - etc.

Digital divides · · ·

We need disaggregated estimates. Why?

- This study applies a Small Area Estimation (SAE) technique that uses census and survey sampling to provide reliable estimates of digital statistics at local municipality levels in South Africa.
- In areas or domains where there are only a few or no samples, SAE aims to provide reliable estimates (Porter et al., 2014).
- The word "small" relates to the sample size of the domain, rather than to the geographical size of the domain, or the size of the domain's population.

Digital divides · · ·

Administrative divisions of South Africa

South Africa

Nine provinces

Metropolitan and district municipalities

Local municipalities $\Longrightarrow$ small areas in this study

Wards

Digital divides · · ·

We need disaggregated estimates. Why?

- SAE techniques impact our everyday life as they often serve as the data basis for political decision-making. For example, the World Bank uses them to develop poverty mapping in both developing and developed countries. The U.S. Census Bureau's uses SAE techniques to develop income and poverty estimates, school age children under poverty, and so on.
- Global frameworks of indicators for monitoring the SDGs, for example, recommend that information be disaggregated not only geographically (in subregions of interest, such as provinces, municipalities, or districts) but also by income group, gender, age, race, ethnicity, immigration status, and disability status.
- It is generally impossible to achieve the desired levels of disaggregation with reasonable accuracy because the reliability of inferences drawn from the indicators decreases as the sample size decreases (Molina, 2022; Hossain et al., 2020; ADB, 2021).

Digital divides · · ·

We need disaggregated estimates. Why?

- When it comes to South Africa, the spatial/administrative unit closest to communities for the location of many basic services and infrastructure is the local municipalities and local wards (Udjo, 2014).
- Thus, providing information at this level is a key consideration for service provision and planning, and targeting of efforts to reduce poverty and inequality, digital inequality, food insecurity, HIV prevalence, malnutrition, under five mortality, and so on.
- The availability of quality information on digital technologies that includes innovations like the Internet, cell phone and social networking platforms at the sub-national level helps:
  - assist businesses to become more productive;
  - creating employment opportunities,
  - for the government and policymakers to provide better public services (World Bank, 2016)
  - and so on

Methodology

Methodology

The FH area-level model

- This model deals with area level summary data and not unit-level data (Ghosh and Rao, 1994; Molina and Rao, 2015).
- This model was first proposed by Fay and Herriot (1979) to estimate the per-capita income in small places of United States of America.
- The problem of interest: $\theta_i = \mathbf{x}_i'\boldsymbol{\beta} + v_i,\ i = 1, \cdots, m$.
- The standard FH model links the sampling model to the population model as follows:

$$\begin{cases} y_i = \mathbf{x}_i'\boldsymbol{\beta} + v_i + e_i, \quad i = 1, ..., m \\ v_i \overset{iid}{\sim} N(0, \sigma_v^2), \\ e_i \overset{ind}{\sim} N(0, D_i) \end{cases} \tag{1}$$

The FH area-level model ...

- There are various ways of estimating the unknown model parameters such as the:
  - Method of moments obtained by Fay and Herriot (1979); and Prasad and Rao (1990)
  - ML and REML proposed by Datta and Lahiri (2000)
- When the unknown parameters in Eqn. (1) are replaced by their estimators, then the EBLUP FH for each small area i is:

$$\hat{\theta}_i^{\mathrm{EB}} = y_i - \hat{\gamma}_i(y_i - \mathbf{x_i'}\hat{\boldsymbol{\beta}}), \quad i = 1, ..., m \tag{2}$$

where

$\hat{\boldsymbol{\beta}} = \left\{\sum_{i=1} \frac{\mathbf{x_i x_i'}}{D_i+\hat{\sigma}_v^2}\right\}^{-1} \left\{\sum_{i=1} \frac{\mathbf{x_i} y_i}{D_i+\hat{\sigma}_v^2}\right\}$ is the WLS estimator of the vector of $\boldsymbol{\beta}$,

$\hat{\gamma}_i = \frac{D_i}{D_i+\hat{\sigma}_v^2}$, $i = 1, \cdots, m$ is the shrinkage factor

The spatial FH model

- The FH model given in Eqn. (1) assumes that the neighbouring small area estimates are spatially uncorrelated.
- In many applications, however, the random effects between the neighbouring small areas are correlated (Rao and Molina, 2015).
- Salvati (2004) proposed extensions of the FH model that allow spatial association instead of independence between neighbouring small areas.
- This process is defined as (Anselin, 1992):

$$v = \rho W v + u \implies v = (I - \rho W)^{-1} u, \tag{3}$$

where ρ is the spatial autoregressive coefficient, $\mathbf{W} = \{w_{ij}\}_{m \times m}$ is a spatial proximity matrix that indicates whether the small areas are neighbours or not $(i, j = 1, \cdots, m)$.

- The spatial EBLUP (EBLUP SFH here after) estimator of θ_i is given by

$$\hat{\theta}_i^{SEB} = x_i \hat{\beta} + b_i' \hat{G} \hat{\mathbf{\Sigma}}^{-1} (y_i - x_i' \hat{\beta}), \tag{4}$$

Mean squared error

- The most common practical problem in small area estimation is measuring the variability associated with the EBLUP FH and EBLUP SFH.
- The variability under these estimators was measured using mean squared error (MSE).
- For example, for the REML estimator, an unbiased analytical estimator of the MSE is:

$$\widehat{\mathrm{MSE}}(\hat{\theta}_i^{\mathrm{EB}}) \approx g_{1i}(\hat{\sigma}_v^2) + g_{2i}(\hat{\sigma}_v^2) + 2g_{3i}(\hat{\sigma}_v^2). \tag{5}$$

- Details about the specification of g components can be found in (Datta and Lahiri, 2000; Chandra et al., 2007; Petrucci et al., 2005; Pratesi and Salvati, 2009).

Data sources

The 2016 Community Survey (2016 CS)

- The **target variables** were drawn from the CS2016 data conducted by Statistics South Africa (Stats SA). The main goal of this survey is to provide indicators such as population count, fertility, mortality, migration, employment, unemployment, the extent of poverty in households, ...

The 2011 Population Census (2011 PC)

- The **auxiliary (covariates) variables** were generated from the 2011 South African Census.
- Around 40 auxiliary variables were selected from the 2011 PC to consider for the modelling.
- The term 'auxiliary information' refers to data sources that have a wider coverage through a large sample size: population census and administrative data are examples of auxiliary data.

Data sources

Auxiliary variables:

These are gender (gen1: male, gen2: female), race (race1: Black African, race2: Coloured, race3: Indian or Asian, race4: White), age (age1: 0-14, age2: 15-24, age3: 25-34, age4: 35-44, age5: 45-54, age6: 55-64 and age7: 65 and above), employment status (emp1: employed, emp2: unemployed, emp3: not economically active, emp4: head employed, emp5: head unemployed and emp6: head not economically active), Employment sector (sec1: in the formal sector, sec2: in the informal sector and sec3: private household), marital status (mar1: married, mar2: living together like married partners, mar3: never married, mar4: widower/widow, mar5: separated, mar6: divorced), education (edu1: no schooling, edu2: some primary, edu3: completed primary, edu4: some secondary, edu5: grade 12, edu6: tertiary, edu7: other), urban area: urban, farm area: farmarea, income (inc1: No income, inc2: R 1 - R 76 800, inc3: R 76 801 - R 614 400, and inc4: R 614 401 or more).

Data sources ...

Out of these auxiliary variables, suitable covariates were chosen using correlation analysis followed by step-wise regression analysis (Chandra et al., 2010). Finally, the following variables were identified for further analysis (see Table 2).

Table 2: Variables for further analysis

Response variables (proportions)	covariates
Male internet users	age2, race1, race2, mar2, mar5, edu1, edu4, edu6, emp4,
Female internet users	age2, race1, mar2, mar5, edu1, edu4, emp4, urban, inc3
Male cell phone owners	gen1, age2, age5, race1, mar6, emp1, emp4, urban, inc2,
Female cell phone owners	gen2, age2, age5, race1, mar6, emp1, emp4, urban, inc3
Male Tablet owners	gen1, race1, edu3, emp1, emp4, farm
Female Tablet owners	gen2, race1, mar6, edu3, emp1, emp4, farm

Digital divides

Results

Results

- The summary of sex-disaggregated estimates on Internet access and use, cell phone and Tablet ownership are summarized in Table 3.
- The EBLUP SFH estimates of internet use by women range from
  - a minimum value equal to 9.69% observed for the local municipality of Renosterberg in the Northern Cape province
  - a maximum equal to 70.46% observed for the local municipality of Mbhashe in the Eastern Cape province.
- The median value of women internet users was equal to 39.53%, therefore 50% of sampled local municipalities (about 107) show an estimated internet use greater or equal to 39.53%.
- The average percentage of women internet users is 40.85%.
- A similar interpretation can be given to men internet users, cell phone owners, male and female Tablet owners.

Results ...

Table 3: Summary of sex-disaggregated estimates on Internet access and use, cell phone and Tablet ownership

	Tablet		*Internet*		*Cellphone*	
	Male	Female	Male	Female	Male	Female
Min.	30.03	15.09	30.09	9.68	26.52	28.81
P25	48.01	38.8	51.85	32.58	42.62	45.1
Median	55.59	44.83	61.02	39.53	48.84	51.29
Mean	55.08	45.34	59.74	40.85	49.03	51.1
Std Dev	9.83	9.83	11.87	11.84	8.16	8.15
P75	61.64	52.41	67.93	48.7	55.02	57.5
Max.	85.4	70.5	91.06	70.46	71.36	73.58

Testing digital inequality between men and women

Descriptive statistics of difference between men's and women's access to and use of digital technologies are reported in Table 4.

- The mean difference of SEBLUP estimates for internet users is 18.90% and the median is 21.00%; and the standard deviation is 23.60%.
- The minimum difference of SEBLUP estimates for internet users is -40.37% and the largest difference of SEBLUP estimates for internet users is 81.38%.

Testing digital inequality ...

Table 4: Summary of differences of digital statistics between men and women

	Tablet owner	Internet acess and use	Cellphone owner
Min	-40.47	-40.37	-47.05
P25	-4.33	3.31	-14.86
Median	10.73	21.00	-2.48
Mean	9.73	18.90	-2.07
Std Dev	19.57	23.60	16.23
P75	22.78	35.28	9.87
Max	70.31	81.38	42.55

Testing digital inequality ...

- Figure 1 shows the SEBLUP estimates for (a) internet users, (b) cell phone owners and (c) Tablet ownership.
- This figure show that there is a very clear difference between men and women internet users in South Africa.
- A similar interpretation can also be given for Tablet and cellphone owners.

Testing digital inequality ...

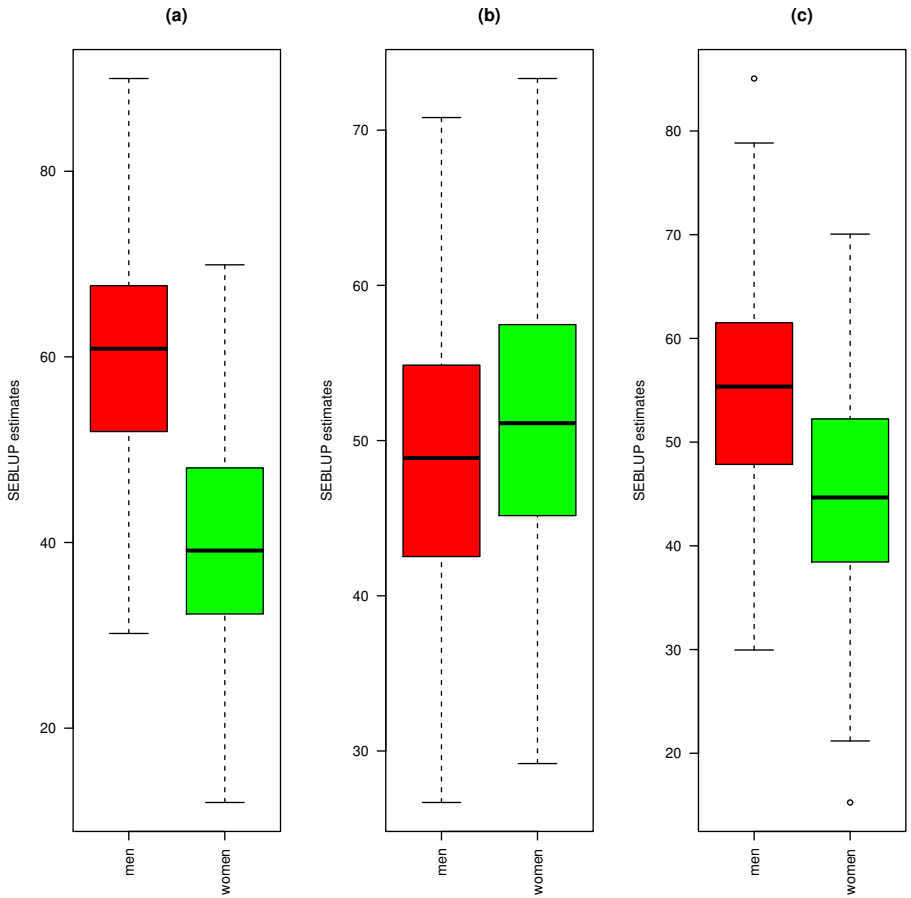

Fig. 1. EBLUP SFH estimates for (a) Internet use, (b) Cell phone ownership and (c) Tablet ownership

Testing digital inequality ...

- From the two sample t test we can conclude that there is gender gaps and differences in access to and use of digital technologies in South Africa.
- The results suggest that when it comes to internet use in South Africa, gender does matter.

Table 5: Two sample t test for EBLUP SFH estimates

Difference	t value	p-value	95\% CI
Internet use: Men vs Women	74.15	0.000	(58.15, 61.33)
Cellphone ownership: Men vs Women	-2.64	0.009	(-3.61, -0.53)
Tablet ownership: Men vs Women	10.32	0.000	(7.89, 11.59)

Residual diagnostics

- Figure 2 presents the distribution of standardized residuals (left), histograms of standardized residuals (centre) and standard normal q-q plots of standardized residuals (right).
- The left plots show that standardized residuals appear to be randomly distributed around zero.
- The centre plots show that histograms of these residuals appear to be symmetric.
- The right plots show that the q-q plots of these residuals are well-distributed along a straight line.
- These plots provide evidence in support of the normality assumption of the standardized residuals.

Residual diagnostics ...

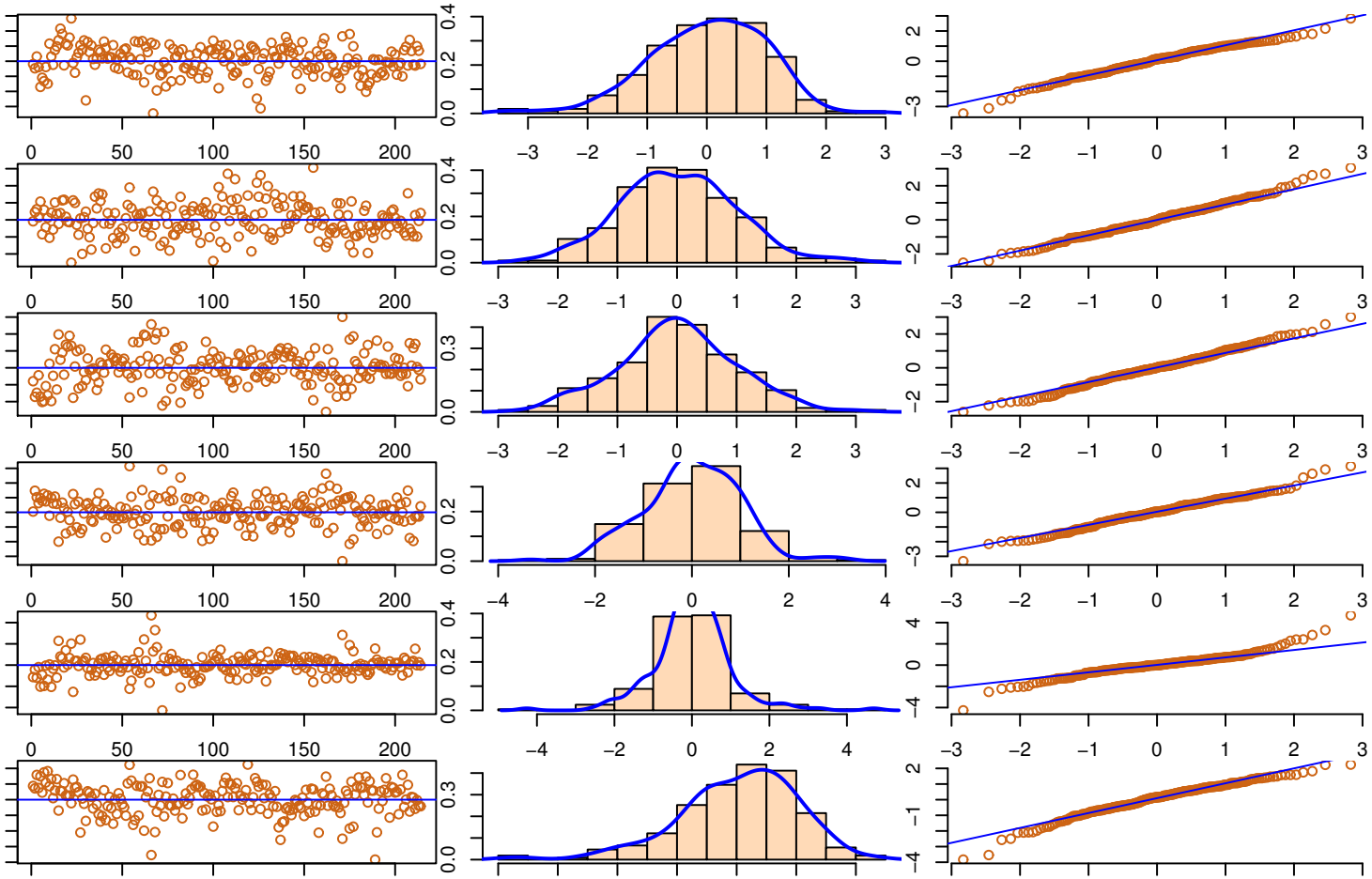

Fig. 2. Panel 1 (women internet users), panel 2 (men internet users), panel 3 (women cellphone owners), panel 4 (men cellphone owners), panel 5 (women tablet owners) and panel 6 (men tablet owners)] (top to bottom).

Map of digital inequality

- Fig. 3 shows the spatial mapping of local municipality-wise digital inequality produced using the log-transformed spatial FH models.
- These maps give a visual illustration of estimated local municipality-wise digital inequalities in South Africa.
- They provide an important source of information on the distribution of local municipality estimates of digital inequality with **yellow colour** corresponding to a better situation (**low digital inequality**), while **red colour** corresponds with a **higher digital inequality**.
- The Western Cape and Gauteng, in contrast, have many local municipalities with low digital inequality.

Map of digital inequality ...

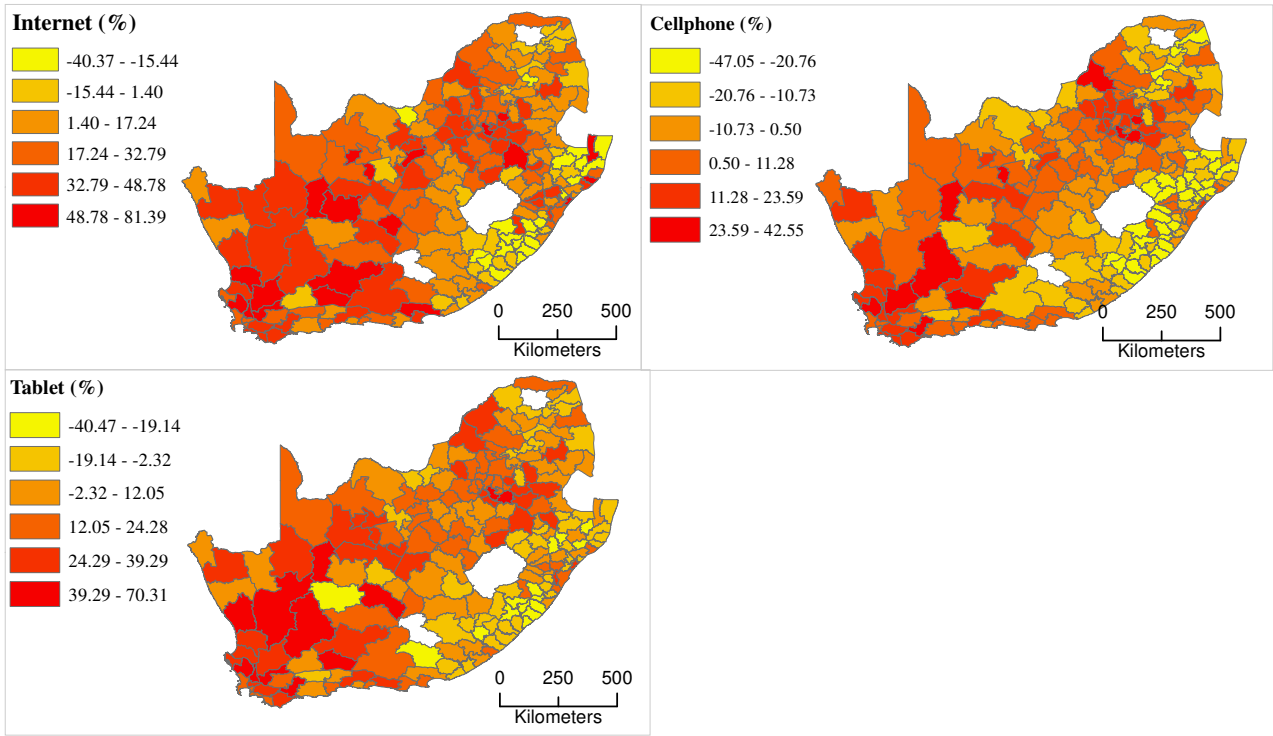

Fig. 3. Map of digital inequality in South Africa

The recommendation areas to bridge the gender gap in ICT access ...

The recommendation areas to bridge the gender gap

- The potential of ICT gains is economic growth, jobs, and services (World Bank, 2016).
- To achieve gender digital equality and improve the livelihood of millions of South Africans, stakeholder groups such as government, non-profit organisations, communities, corporate leaders and individuals should address the followings;
  i) Upgrade women's digital skills, relevance and confidence
  ii) Provide affordable access
  iii) Warn and motivate women to take advantage of the opportunities offered by the digital transformation.
  iv) Innovative plans and actions.

Conclusion and policy implications

Conclusion ...

- According to the results (Tables and Figures), in South Africa, there are gender inequality concerning access to and use of digital technologies.
- In particular, gender matters regarding internet use in South Africa.
- For example, internet use by women ranges from
  - a minimum value equal to 9.69% observed for the local municipality of Renosterberg in the Northern Cape province
    to
  - a maximum equal to 70.46% observed for the local municipality of Mbhashe in the Eastern Cape province.
- This shows that the analysis at the national level does mask important dissimilarities that characterize these areas.

Conclusion and policy implications ...

Policy implications:

- **Allocate funds:** constant monitoring of socio-economic indicators such as digital inequality at the lower geographical level is needed to adequately allocate funds and to assess the effectiveness of their spending (Pratesi, 2015).
- With the help of the map, policy decisions could be made, resources allocated, and measures to eradicate digital inequality could be planned and evaluated.
- Government agencies, non-profit organizations, communities, corporate leaders, and individuals in South Africa will be able to work together to allocate budgets and intervene to bridge gender digital inequalities based on the disaggregated information from this study.

Conclusion and policy implications ...

Prioritize and allocate resources:

- The SAE methods helped us to identify the marginalized local municipalities.
- In other words, the national government, as well as municipality level officials, have the information to prioritize and allocate resources.

Conclusion and policy implications ...

Limitations and future work:

- In this study, the data sources used do not reflect the current reality in South Africa.
- In the South African context, the only nationally representative data available within the coming 2 to 3 years will be the 2022 Population Census.
- Based on previous experience and the experience of other countries, the Census 2022 report will probably be released in 2024, and we will have the most up-to-date data on digital statistics at the local level.
- Therefore, future research should focus on the Census 2022 data and compare it to any previous study to determine differences in the digital gender divide.

Some important references

- Benavent R. and Morales D. (2016). Multivariate Fay–Herriot models for small area estimation. *Computational Statistics and Data Analysis*, 94, 372-390.
- Rao, J.N.K. and Molina, I. *Small Area Estimation*. John Wiley and Sons, Inc., New York, 2015.
- Esteban, M.D., Morales, D., Perez, A. and Santamaria, L. (2012) Small area estimation of poverty proportions under area-level time models. *Computational Statistics and Data Analysis*, 56, 2840-2855.

Thank you very much for your attention!

Questions ?

Acknowledgement:

Thank you to ERSA for organizing my travel and accommodation.