

Facebook data validity analysis for studying immigration to Mexico

Tania Varona, El Colegio de México

Claudia Masferrer, El Colegio de México

Victoria Prieto, Universidad de la República

Martín Pedemonte, Universidad de la República



Background



International migration is the most complex demographic component to measure

It is hard to do it in a regular and timely manner. Its most precise estimate is often available every ten years using census data.

Facebook

Facebook has recently emerged as an alternative source for alleviating the lack of up-to-date and detailed information

We can describe stocks and characteristics of migrant populations more frequently.

Background



There has been a surge of research that analyze international migration incorporating the Facebook data definition of *expat*.

For example, studies have used the expatriate variable for nowcasting immigration in Germany (Alexander et al., 2020).

To estimate Venezuelan immigrants in Colombia (Palotti et al., 2020).



To assess the quality of Facebook data to measure immigration stocks by contrasting it with traditional sources of information such as the American Community Survey in the United States or the Labor Force Survey in the United Kingdom (Zagheni et al., 2017; Spyrtatos et al., 2019).

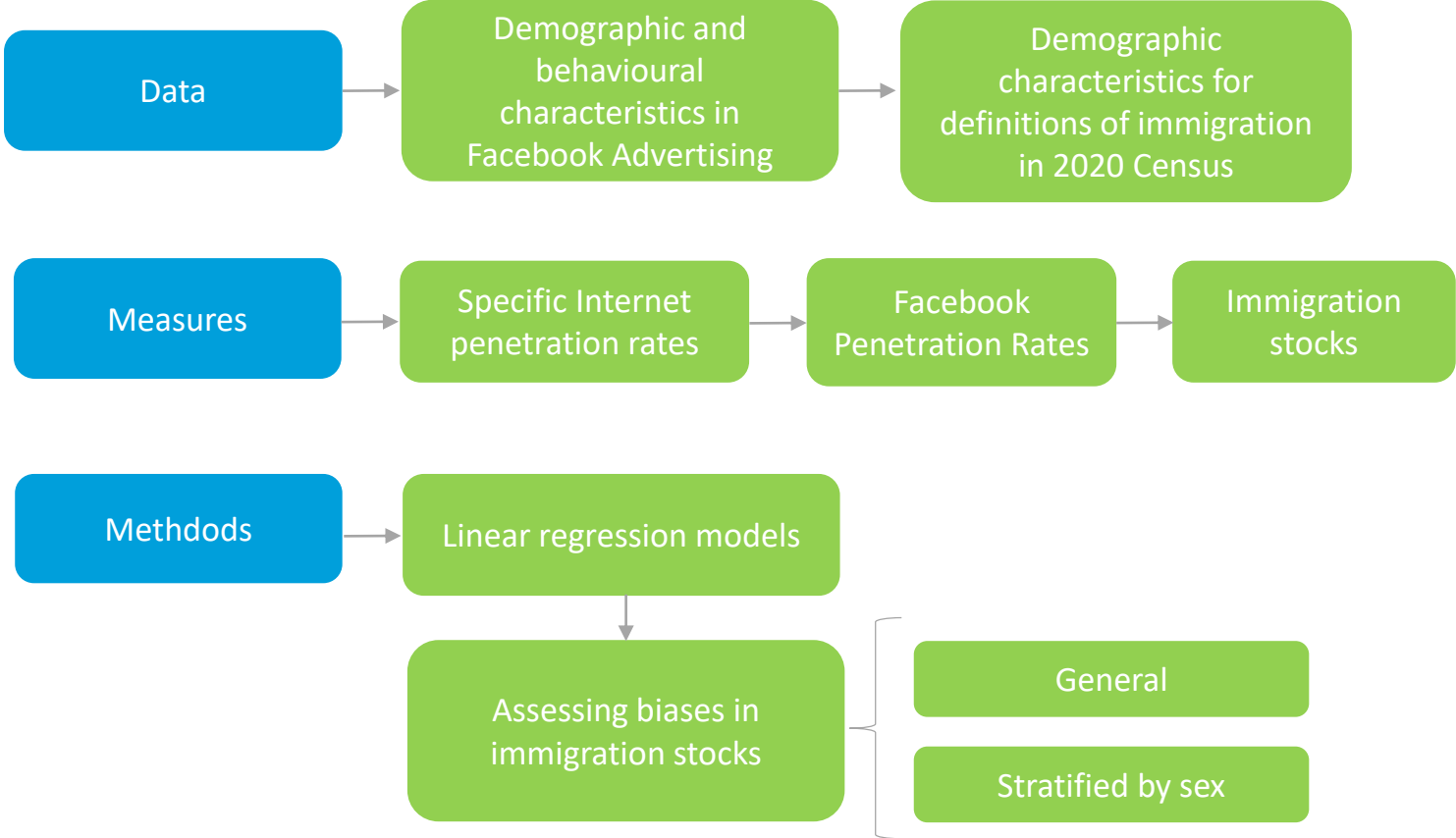
Motivation and aims

Facebook data have not been evaluated with a population census. Motivated by informing migration processes in Latin America, we compare [Facebook data](#) with data from the [2020 Mexican Census](#).

Specifically, this study aims to:

- Identify which definition of immigration is closest to the Facebook expat variable.
- Understand the direction of biases in demographic variables: sex, age, and country of origin.
- Propose a function to estimate immigration stocks in intercensal periods.

Data, measures, and methods



Data

3-group age

13 Latin American countries

6 definitions of immigration

Variable	Census	Facebook
Sex	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Age	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Country of origin	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Immigrants by country of birth	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Immigrants by previous residence in 2015	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Recent immigrants by place of birth	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Recent immigrants residing in 2015 in the same country where they were born	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Recent immigrants residing in 2015 in a different country from their country of birth	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Mexican returnees residing abroad in 2015	<input checked="" type="checkbox"/>	<input type="checkbox"/>

2020 Mexican Population Census

- Microdata

2020 Facebook Advertising data

- Max MAU /month – expats

Variables

- Sex: men and women
- Age: 13–24; 25–54; and 55–65 years
- 13 Latin Americans countries: Argentina, Brazil, Chile, Colombia, Cuba, El Salvador, Guatemala, Haiti, Honduras, Nicaragua, Peru, Dominican Republic, y Venezuela

Measures

Internet penetration rates

- Identify differences in internet penetration rates according to the age, gender, and country of birth or previous residence of immigrants
- Indicator calculated using 2020 Census data

Facebook Penetration Rate

- % Facebook users in Mexico (migrants and non-migrants)
- Indicator calculated using Facebook and Census data

% Immigrants/expats in total population/Facebook users

- Independent and dependent for modelling

Methods

Graphical assessment

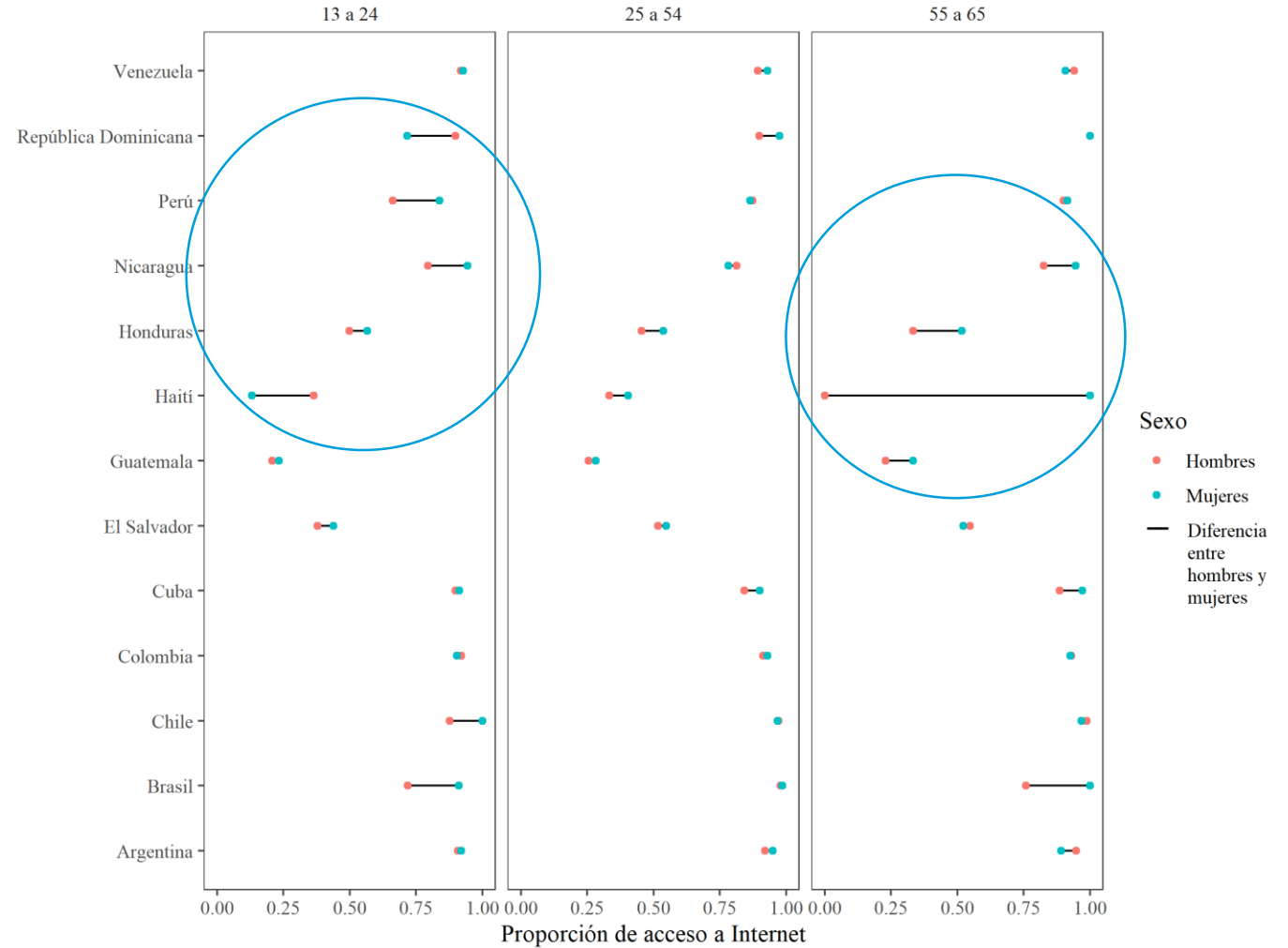
- Descriptive analysis of the differences in Internet penetration rates and biases according to demographic variables.

Linear regression models

- 24 models (4 models x 6 immigration definitions)
- 36 models segregated by sex (3 models x 6 immigration definitions x 2 sexes)

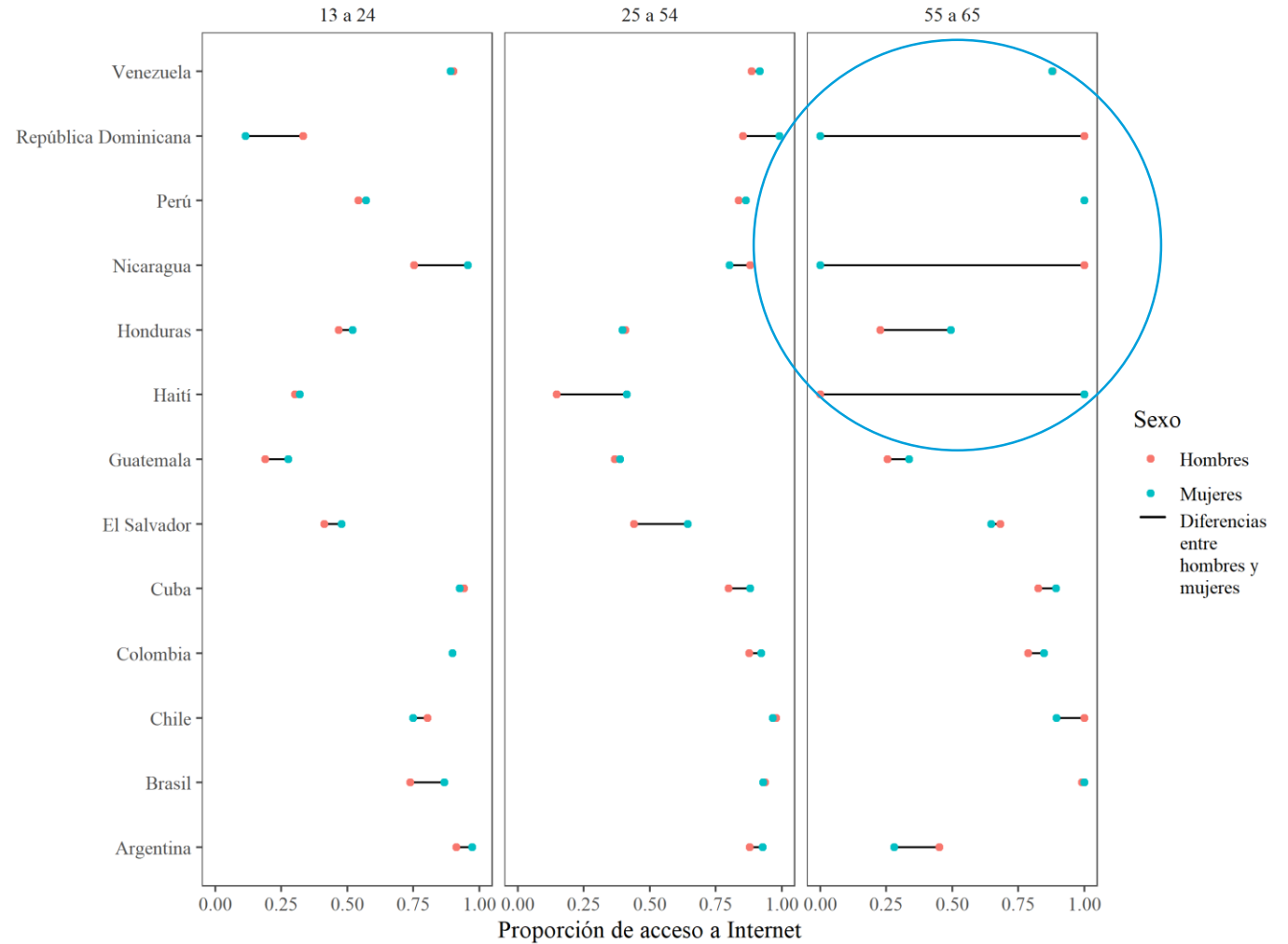
Internet penetration rates

Graph 1.
Sex and age specific
Internet penetration rates
by country of birth



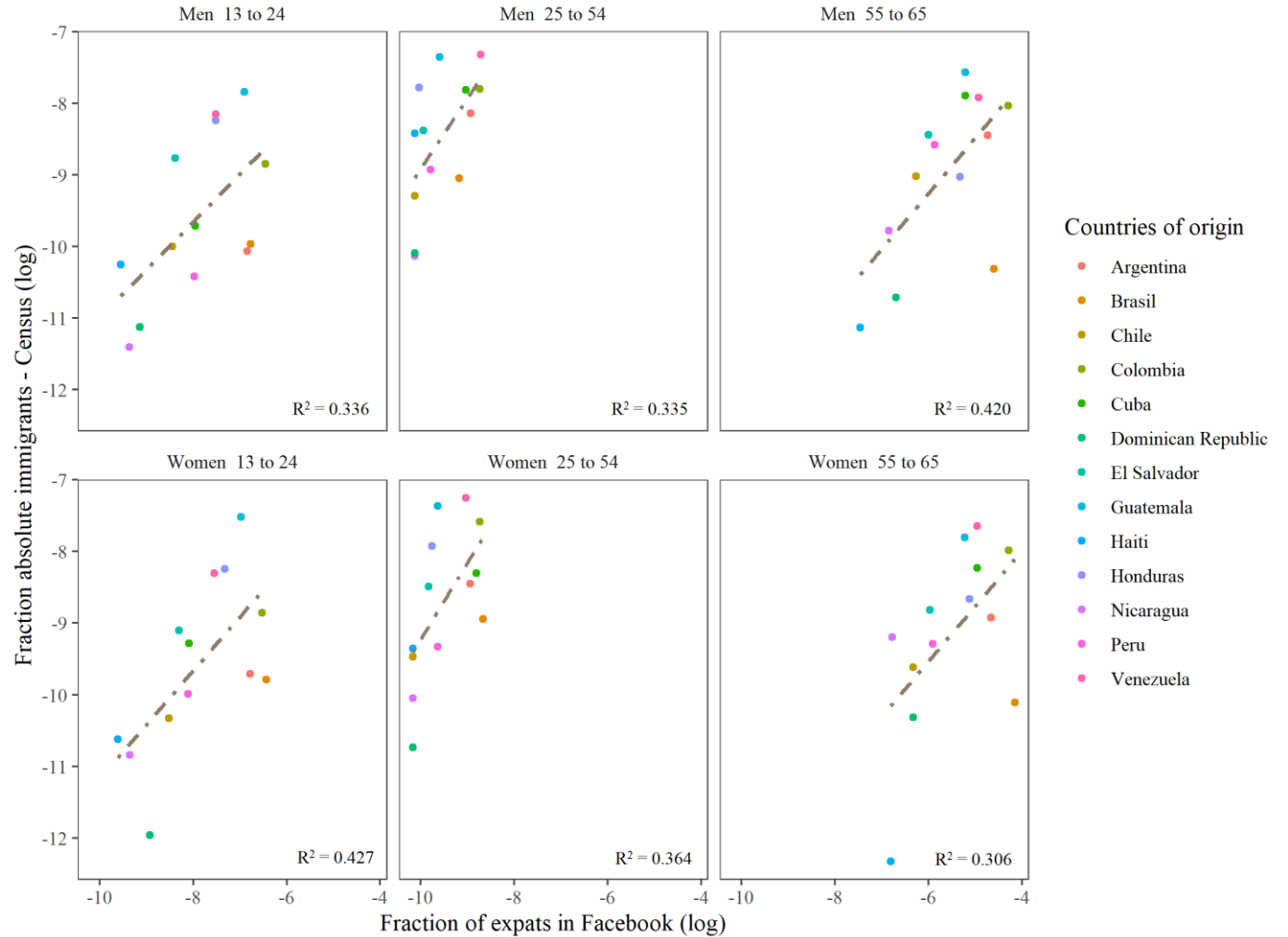
Internet penetration rates

Graph 2.
Sex and age specific
Internet penetration rates
by country of previous
residence 2015



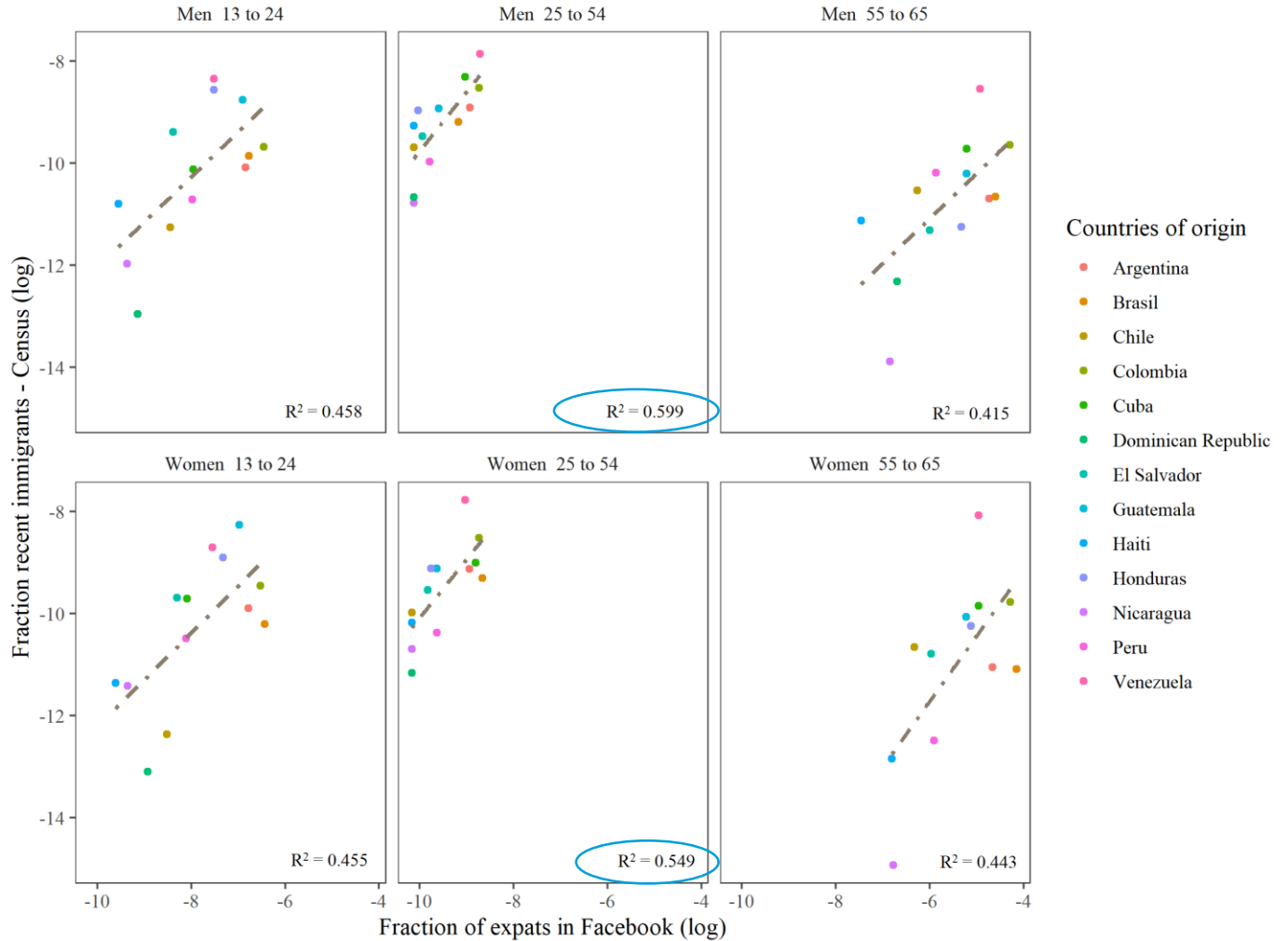
Results, Graphical assessment

Graph 3. Correlation between the proportion of immigrants in the total population by country of birth and the proportion of expats users in the total Facebook MAU



Results, Graphical assessment

Graph 3. Correlation between the proportion of immigrants in the total population by country of previous residence 2015 and the proportion of expats users in the total Facebook MAU



Results, Analytical assessment

Linear regression models, previous residency in 2015

- The coefficient of expats is significant, suggesting a correlation with this definition of immigration.
- Negative coefficients on sex and age indicate an overestimation of expats relative to recent immigrants.
- Positive and significant coefficients on country of origin indicate an underestimation.
- With this analysis, it is feasible to obtain a linear function to estimate immigration in Mexico in intercensal periods with the Facebook data once the biases are considered.

	Immigration stocks by country of previous residence 2015 (log)			
	(1)	(2)	(3)	(4)
Expat stocks (log)	-0.073 (0.087)	-0.077 (0.087)	1.016*** (0.129)	0.712*** (0.231)
Woman		-0.187 (0.322)	0.013 (0.217)	-0.085 (0.150)
25 to 54			2.505*** (0.338)	2.004*** (0.417)
55 to 65			-3.056*** (0.404)	-2.392*** (0.567)
Brasil				-0.216 (0.359)
Chile				0.277 (0.496)
Colombia				0.476 (0.364)
Cuba				0.885** (0.377)
El Salvador				0.825* (0.460)
Guatemala				1.055*** (0.371)
Haiti				0.555 (0.610)
Honduras				0.953** (0.392)
Nicaragua				-0.923 (0.576)
Peru				0.020 (0.434)
Dominican Republic				-1.046* (0.543)
Venezuela				1.960*** (0.363)
Constant	-10.727*** (0.684)	-10.474*** (0.814)	-2.227** (1.048)	-4.856*** (1.590)
Observations	77	77	77	77
R ²	0.009	0.014	0.567	0.846
Adj. R ²	-0.004	-0.013	0.543	0.806
AIC	274.7123	276.3619	217.0173	161.1651
BIC	281.7437	285.7371	231.0801	203.3536
Residual Std. Error	1.404 (df= 75)	1.410 (df= 74)	0.947 (df= 72)	0.618 (df= 60)
F Statistic	0.717 (df= 1; 75)	0.524 (df= 2; 74)	23.558*** (df= 4; 72)	20.673*** (df= 16; 60)

Note: *p<0.1; **p<0.05; ***p<0.01; parenthesis indicates standard errors.

Source: Facebook 2020 and Mexican Census 2020.

Results, Analytical assessment

Linear regression models, previous residency in 2015

The model for males has greater explanatory power than the model for females (adj. R2 = 79.1 vs. adj. R2 = 76.1).

	Immigration stocks by country of previous residence 2015 (log)					
	Men			Women		
	(1)	(2)	(3)	(4)	(5)	(6)
Expat stocks (log)	-0.109 (0.113)	0.888*** (0.176)	0.668* (0.347)	-0.041 (0.136)	1.153*** (0.193)	0.768* (0.388)
Woman		2.394*** (0.464)	2.029*** (0.623)		2.626*** (0.503)	1.997*** (0.695)
25 to 54		-2.741*** (0.559)	-2.206** (0.877)		-3.368*** (0.598)	-2.610*** (0.913)
55 to 65			-0.255 (0.510)			-0.160 (0.596)
Brasil			0.434 (0.728)			0.133 (0.818)
Chile			0.427 (0.503)			0.518 (0.610)
Colombia			0.841 (0.522)			0.938 (0.635)
Cuba			0.667 (0.654)			0.999 (0.774)
El Salvador			0.922* (0.521)			1.189* (0.616)
Guatemala			0.881 (0.871)			0.261 (1.044)
Haiti			0.711 (0.536)			1.213* (0.670)
Honduras			-0.998 (0.844)			-0.826 (0.961)
Nicaragua			0.336 (0.622)			-0.286 (0.721)
Peru			-0.966 (0.763)			-1.173 (0.959)
Dominican Republic			1.905*** (0.511)			2.009*** (0.602)
Constant	-10.904*** (0.878)	-3.255** (1.396)	-5.352** (2.452)	-10.569*** (1.088)	-1.080 (1.569)	-4.521 (2.795)
Observations	39	39	39	38	38	38
R ²	0.025	0.552	0.873	0.003	0.589	0.858
Adjusted R ²	-0.002	0.514	0.791	-0.025	0.553	0.761
AIC	136.39075	110.02807	84.76393	143.13631	113.39840	97.13743
BIC	141.3814	118.3459	113.0445	148.0491	121.5863	124.9764
Residual Std. Error	1.322 (df = 37)	0.921 (df = 35)	0.604 (df = 23)	1.511 (df = 36)	0.997 (df = 34)	0.730 (df = 22)
F Statistic	0.935 (df = 1; 37)	14.389*** (df = 3; 35)	10.578*** (df = 15; 23)	0.091 (df = 1; 36)	16.275*** (df = 3; 34)	8.841*** (df = 15; 22)

Nota:

*p<0.1; **p<0.05; ***p<0.01

Conclusions

- There are differences in internet access among the migrant population. Overall, women and migrants by country of birth have greater internet access than migrants by country of previous residence.
- Using the 2020 census data allowed us to test several definitions of immigration and to use specific immigration stocks differentiated by age, sex and country of origin to understand which definition is closest to the expatriate data.
- We find that the closest definition is that of immigration by previous residence in 2015.
- We identified overestimation biases in sex and age, as well as underestimation in some country-of-origin coefficients.
- We also identify overestimation biases by sex, age and countries of origin.
- For Mexico, the model for males has greater explanatory power than the model for females (adj. R2 = 79.1 vs. adj. R2 = 76.1).
- Findings suggest alternative measures of immigration stocks in real time for intercensal periods using Facebook MAU data correcting for specific biases on age, sex and country of origin.

Thanks!

tvarona@colmex.mx