30th International Population Conference
13-18 July 2025
Brisbane, Australia

Quantifying the Intraregional Migration in Latin America Using Facebook: a Validity Assessment based on Household Surveys

Abstract

Digital traces are increasingly being considered as alternative data sources for monitoring international displacement in real time, although evidence of their validity is still limited in low- and middle-income regions. While web-based social media data, such as Facebook users' information, have been used to study migration trends in Latin America, household surveys have not yet been used to validate these data. This paper examines how Facebook users identified as having 'previously lived in' specific locations match migrant population figures reported by household surveys in Latin American countries for 2021, 2022 and 2023. Multivariate analysis shows a high correlation between the two datasets, albeit with a general bias towards overestimating of Facebook by origin and destination. Although Facebook data is more consistent with UN data than with survey data, the results still show a good fit with traditional sources, confirming the value of using Facebook data for real-time estimation of migrant stocks.

Key words: International Migration, Big Data, Digital and Computational Demography

Introduction

The measurement and characterization of migration stocks in Latin America has mostly relied on statistical data sources, with a particular dominance of census data (Martínez-Pizarro & Rivera-Orrego, 2016; Calvelo, 2011; Cerrutti, 2009). However, censuses have their own limitations, as do demographic estimates and household surveys, all of which are affected by issues of timeliness (Perdomo Rico, 2022; Gutiérrez et al., 2020; Martínez Pizarro, 2009). Therefore, researchers need to compare and complement them with alternative data to adequately capture changes in the size and composition of international migration in real time.

Digital trace data have been used to track migrant stocks in real time in Western countries (Rampazzo et al., 2021; Alexander, Polimis & Zagheni, 2020; Zagheni, Weber & Gummadi, 2017), but less is known about their validity for addressing international migration in the Latin American context. To the best of our knowledge, the few validation exercises focused on Latin America have been carried out using the 2020 Mexican Population Census, United Nations (UN) estimates and information from the R4V Platform (Varona et al., 2024; Prieto Rosas et al., 2022; Palotti et al., 2020). However, despite their annual availability, household survey data have not yet been used for this purpose.

This work aims to contribute to the validation of Facebook, Instagram and Messenger -hereafter Facebook- data by analyzing the extent to which the number of Facebook users tagged as 'users who previously lived in a country other than their current country of residence' approximates the number of absolute migrants from household surveys. To this end, we compare data from Facebook with household surveys using multivariate linear models by sex, country of origin and destination for eight Latin American countries.

Background

The use of digital trace data for demographic research is a recent development (Salganik, 2018; Cesare et al., 2018). It has been used to study the three demographic domains (mortality, fertility and migration) (Basellini et al., 2021; Wilde, Chen & Lohmann, 2020; Grow et al., 2020; Alexander, Polimis & Zagheni, 2020; Rampazzo et al., 2018; Zagheni, Weber & Gummadi, 2017; Billari, D'amuri & Marcucci, 2016). However, it is in the field of international migration that it has proved its worth. International migration researchers have been the leading experts in the use of digital traces. Thus, data from emails (Zagheni & Weber, 2012), Twitter (Gil-Clavel, Grow & Bijlsma, 2023) and Facebook (Rampazzo et al., 2021; Alexander, Polimis & Zagheni, 2020; Hsiao et al., 2020; Spyratos et al., 2019, 2020; Zagheni et al., 2018; Zagheni, Weber & Gummadi, 2017) are used to study different aspects of migration, ranging from estimating migrant stocks to migrants integration.

In the specific case of estimating international migration stocks, Facebook data has been the most used so far (Rampazzo et al., 2021; Alexander, Polimis & Zagheni, 2020; Hsiao et al., 2020; Spyratos et al., 2019, 2020; Zagheni et al., 2018; Zagheni, Weber & Gummadi, 2017). In the case of Latin America, it is important to highlight three papers that suggest that Facebook data on expat users matches quite well with the immigrant stocks reported by UN estimates, census and household survey data. First, Prieto Rosas et al. (2022) found high levels of correlation between Facebook data for 2019-2020 and UN International Migrant Stock (UN

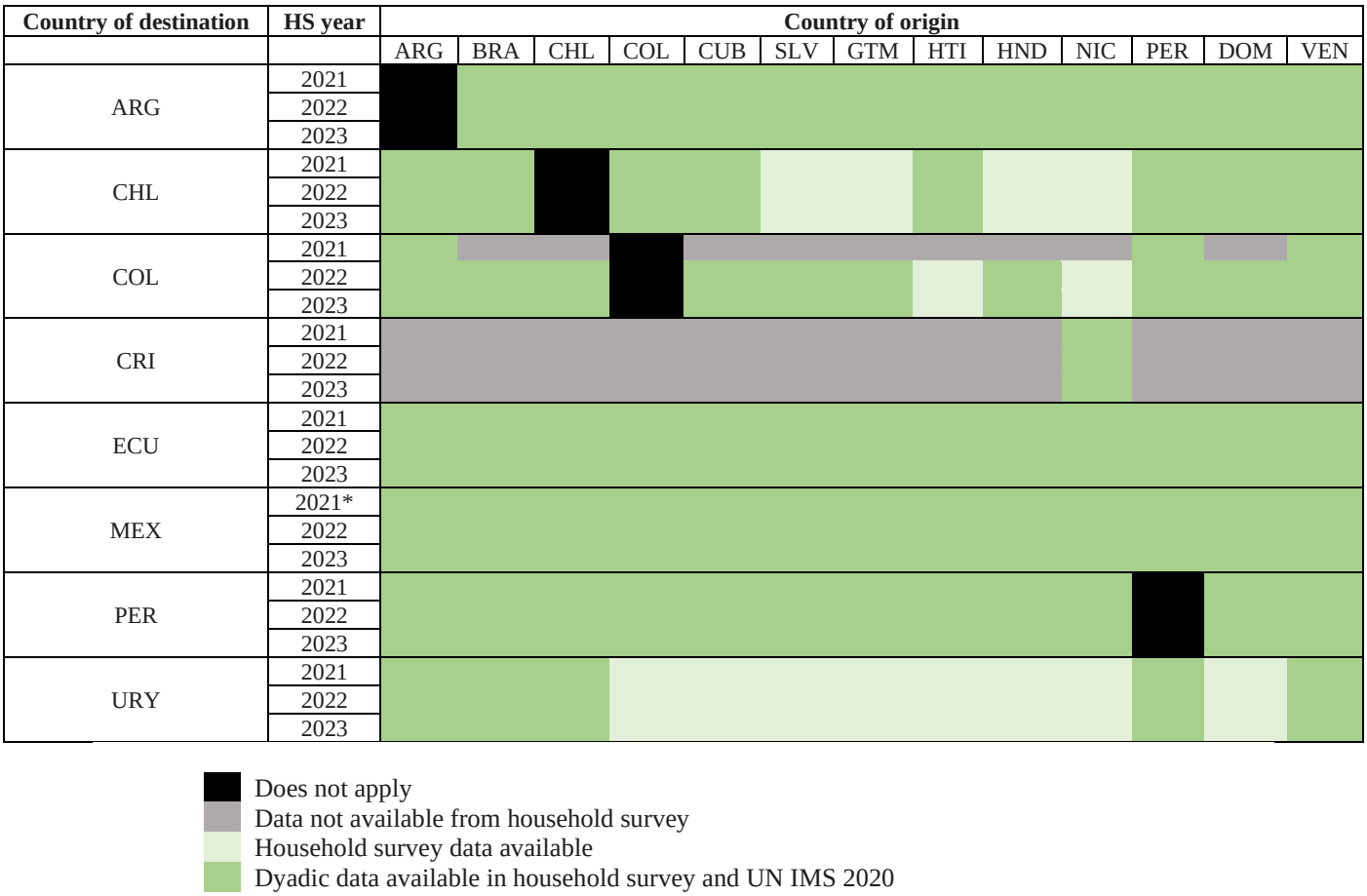
IMS) estimates for 21 destination countries in Latin America and the Caribbean. However, the calibration exercise shows that Facebook tends to overrepresent stocks by destination country and underrepresent by country of origin compared to UN data. Second, using census data, Varona (2022) and Varona et al. (2024) found a similar fit but in this case their results suggest an underrepresentation of foreign origin relative to the 2020 Mexican Population and Housing Census. Third, Rosati, Cerrutti & Maguire (2021), using household surveys but restricting their analysis to Venezuelans in the Autonomous City of Buenos Aires, found that Facebook overrepresents the number of people of this origin and underrepresents those over the age of 24.

Therefore, based on the literature, we do not expect to find significant differences by sex (Prieto Rosas et al., 2022), but we would expect Facebook to generally overestimate the size of the migrant stock by destination country (Rosati, Cerrutti & Maguire, 2021; Prieto Rosas et al., 2022), despite possible variation by country of origin (Varona et al., 2024). We do not expect to find significant differences in this pattern by sex (Prieto Rosas et al., 2022).

Data and methods

We used three data sources for eight Latin American destination countries: Argentina, Chile, Colombia, Costa Rica, Ecuador, Mexico, Peru and Uruguay (Figure 1). The first source is microdata from household surveys conducted in these countries. The second source is Facebook, Instagram and Messenger Daily Active Users (DAU) data, restricted to users who previously resided in Latin American and Caribbean countries and currently reside in the selected destination countries. A synthetic measure of DAU was obtained using a bootstrap procedure by calculating the mean of 2,500 medians for each origin-destination pair, disaggregated by sex and aligned with the household survey periods. Both datasets cover the period 2021-2023. The third source corresponds to the 2020 UN International Migration Stock (IMS) estimates, from which we extracted origin-destination dyads by sex. For all three data sources, we focused on individuals aged 18-65, disaggregated by sex. As age structure information was not available at the dyad level for the UN IMS data, we imputed the age distribution of the total migrant stock in each destination country. We used multivariate linear models to estimate the associations and biases by destination, origin and sex. We estimated a multivariate linear model for each year of the household surveys (2021 to 2023) and for the UN IMS 2020 data.

Figure 1. Description of the dyads analyzed by country of origin, destination, year and source of information



Notes: *For the first two trimesters of 2021, disaggregated information is only available for Guatemala.
Source: based on Facebook data; UN IMS 2020; Encuesta Permanente de Hogares (Argentina), Encuesta Nacional de Empleo (Chile), Gran Encuesta Integrada de Hogares (Colombia), Encuesta Nacional de Hogares (Costa Rica), Encuesta Nacional de Empleo, Desempleo y Subempleo (Ecuador), Encuesta Nacional de Ocupación y Empleo (Mexico), Encuesta Nacional de Hogares (Peru) and Encuesta Continua de Hogares (Uruguay).

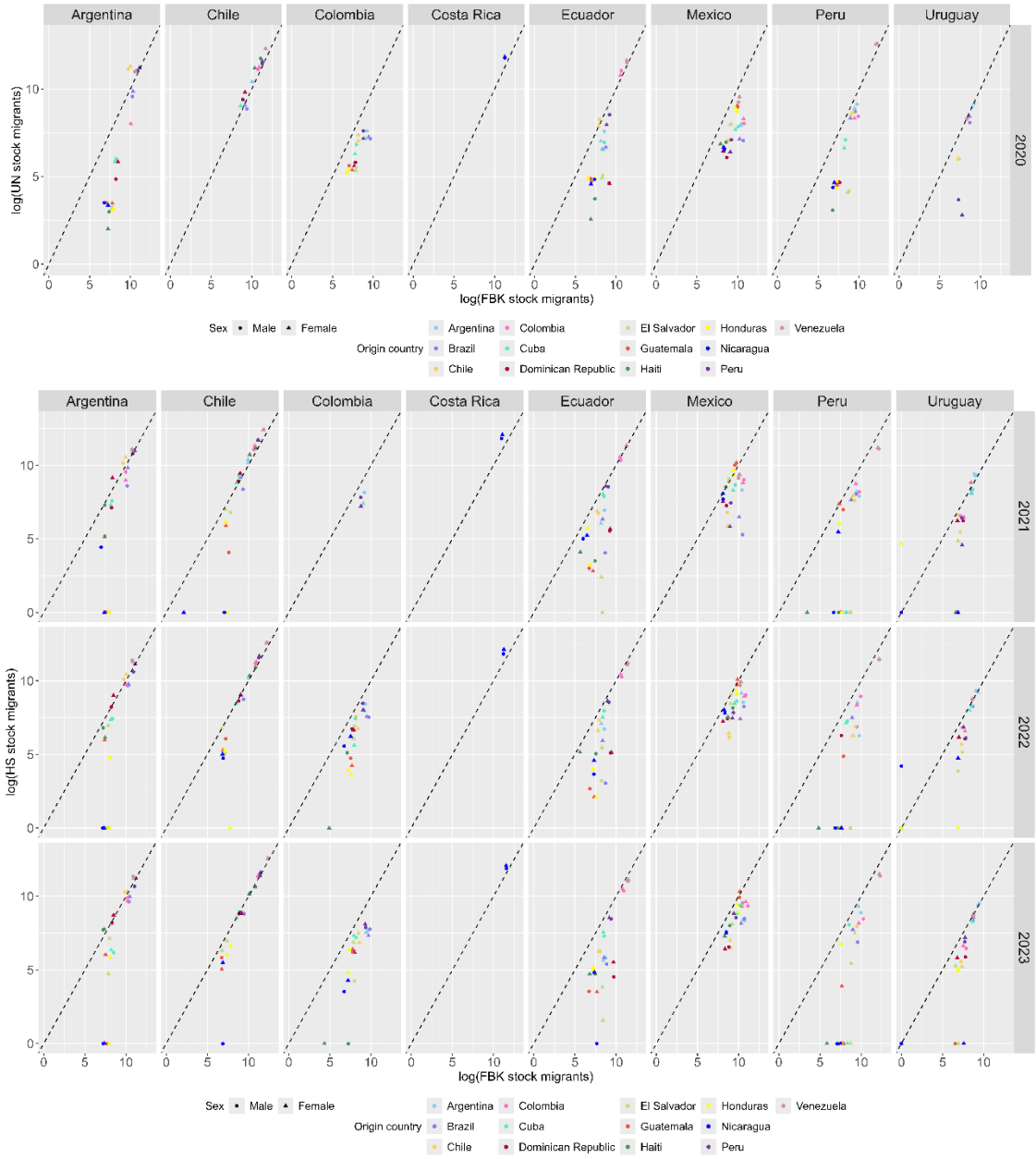
Results

Graphical analysis reveals five key findings (Figure 2). First, Facebook data are generally consistent with both household survey data and UN estimates of migrant stocks. Second, Facebook data tends to overestimate migrant stocks across the board. Third, the closeness to the identity line is stronger for the largest origin communities in each destination, while greater dispersion is observed for smaller migrant populations from Central America and the Caribbean. Fourth, the fit varies across destination countries, with Argentina and Chile showing the most consistent stock magnitudes. Finally, there appear to be no significant sex differences in the data.

The multivariate models reinforce some of these findings (not shown here). First, there is a significant correlation between Facebook and household surveys, with an R-squared close to 0.7 for all years. However, this correlation appears to be stronger between Facebook and UN IMS data. Second, there is a systematic overestimation of stocks by Facebook for all origins and destinations considered. The exception is the case of Costa Rica, where the stock of

Nicaraguan migrants is underestimated. The comparison with the UN shows similar results for the countries of origin in terms of overestimation. Third, the existence of significant differences between Facebook and household surveys by sex is not consistent, and is not significant in the UN estimate.

Figure 2. Ratio of Facebook migrant stock to estimates based on household surveys and UN IMS by country of origin and destination by year. Selected countries, 2020-2023



Note: For better visualization, the data for the Colombia-Venezuela dyad are not shown due to its larger magnitude.

Source: Based on Facebook data; UN IMS 2020; Encuesta Permanente de Hogares (Argentina), Encuesta Nacional de Empleo (Chile), Gran Encuesta Integrada de Hogares (Colombia), Encuesta Nacional de Hogares (Costa Rica), Encuesta Nacional de Empleo, Desempleo y Subempleo (Ecuador), Encuesta Nacional de Ocupación y Empleo (Mexico), Encuesta Nacional de Hogares (Peru) and Encuesta Continua de Hogares (Uruguay).

References

Alexander, M., Polimis, K., & Zagheni, E. (2020). Combining social media and survey data to nowcast migrant stocks in the United States. *Population Research and Policy Review*, *41*, 1-28.

Basellini, U., Alburez-Gutierrez, D., Del Fava, E., Perrotta, D., Bonetti, M., Camarda, C. G., & Zagheni, E. (2021). Linking excess mortality to mobility data during the first wave of COVID-19 in England and Wales. *SSM-Population Health*, *14*, 100799.

Billari, F., D'Amuri, F., & Marcucci, J. (2016). Forecasting births using Google. In *Carma 2016: 1st international conference on advanced research methods in analytics* (pp. 119-119), Universitat Politècnica de València.

Calvelo, L. (2011). Viejos y nuevos asuntos en las estimaciones de la migración internacional en América Latina y el Caribe. *Serie Población y Desarrollo CEPAL*, No. 98.

Cerrutti, M. (2009). Gender and Intra-Regional Migration in South America. *Human Development Research Paper UNDP*, No. 12.

Cesare, N., Lee, H., McCormick, T., Spiro, E., & Zagheni, E. (2018). Promises and Pitfalls of Using Digital Traces for Demographic Research. *Demography*, *55*(5), 1979–1999.

Gil-Clavel, S., Grow, A., & Bijlsma, M. J. (2023). Migration Policies and Immigrants' Language Acquisition in EU-15: Evidence from Twitter. *Population and Development Review*, *49*(3), 469-497.

Grow, A., Perrotta, D., Del Fava, E., Cimentada, J., Rampazzo, F., Gil-Clavel, S., & Zagheni, E. (2020). Addressing public health emergencies via Facebook surveys: Advantages, challenges, and practical considerations. *Journal of Medical Internet Research*, *22*(12), e20653.

Gutiérrez, A., Mancero, X., Fuentes, A., López, F., & Molina, F. (2020). Criterios de calidad en la estimación de indicadores a partir de encuestas de hogares: una aplicación a la migración internacional. *Serie Estudios Estadísticos CEPAL*, No. 101.

Hsiao, Y., Fiorio, L., Wakefield, J., & Zagheni, E. (2020). Modeling the bias of digital data: an approach to combining digital and survey data to estimate and predict migration trends. *MPIDR Working Paper*, No. 19.

Martínez Pizarro, J. (2009). Medición e información sobre la migración internacional a partir de los censos: lecciones, desafíos y oportunidades. *Notas de población*, *36*(88), 97-133.

Martínez-Pizarro, J., & Rivera-Orrego, C. (2016). Nuevas tendencias y dinámicas migratorias en América Latina y el Caribe. *Serie Población y Desarrollo*, CEPAL, No. 114.

Palotti, J., Adler, N., Morales-Guzman, A., Villaveces, J., Sekara, V., Herranz, M. G., ..., & Weber, I. (2020). Monitoring of the Venezuelan exodus through Facebook's advertising platform. *PLoS ONE*, *15*(3), e0230455.

Perdomo Rico, J. C. (2022). Visibilizando a los migrantes: revisión de la información sobre migrantes en censos y encuestas de hogares en América Latina y El Caribe. *Nota técnica del Banco Interamericano de Desarrollo*, No. 2.

Prieto Rosas, V., Bengochea, J., Pedemonte, M., & Zagheni, E. (2022). Using Facebook to estimate migrant stock in Latin American and Caribbean countries, *MPIDR Lab in Digital and Computational Demography Workshop.*

Rampazzo, F., Bijak, J., Vitali, A., Weber, I., & Zagheni, E. (2021). A framework for estimating migrant stocks using digital traces and survey data: An application in the United Kingdom. *Demography*, *58*(6), 2193-2218.

Rampazzo, F., Zagheni, E., Weber, I., Testa, M. R., & Billari, F. (2018). Mater Certa Est, Pater Numquam: What Can Facebook Advertising Data Tell Us about Male Fertility Rates? *Twelfth International AAAI Conference on Web and Social Media*.

Rosati, G., Cerruti, M., & Maguire, T. (2021). Un acercamiento a la población venezolana a nivel local mediante datos de Facebook, *XVI Jornadas Argentinas de Estudios de Población y III Congreso Internacional de Población del Cono Sur*.

Salganik, M. J. (2018). *Bit by Bit. Social research in the digital age*, Princeton University Press.

Spyratos, S., Vespe, M., Natale, F., Weber, I., Zagheni, E., & Rango, M. (2019). Quantifying international human mobility patterns using Facebook Network data. *PLoS ONE*, *14*(10), e0224134.

Spyratos, S., Vespe, M., Natale, F., Iacus, S. M., & Santamaria, C. (2020). Explaining the travelling behaviour of migrants using Facebook audience estimates. *PLoS ONE, 15*(9), e0238947.

Varona, T. (2022). Análisis de la validez de datos de Facebook para el estudio de la inmigración en México (Master's Thesis, El Colegio de México). Retrieved from https://repositorio.colmex.mx/concern/theses/sj139396d?f%5Bcenter_sim%5D%5B%5D=Centro+de+Estudios+Demogr%C3%A1ficos%2C+Urbanos+y+Ambientales&f%5Bgeographic_coverage_sim%5D%5B%5D=M%C3%A9xico&locale=es&per_page=10

Varona, T., Masferrer, C., Prieto Rosas, V., & Pedemonte, M. (2024). Which definition of migration better fits Facebook 'expats'? A response using Mexican census data. *Demographic Research*, *50*(39), 1171-1184.

Wilde, J., Chen, W., & Lohmann, S. (2020). COVID-19 and the Future of US Fertility: What Can We Learn from Google? *SocArXiv Papers*. https://doi.org/10.31235/osf.io/2bgqs

Zagheni, E., Polimis, K., Alexander, M., Weber, I., & Billari, F. C. (2018). Combining social media data and traditional surveys to nowcast migration stocks, *Annual Meeting of the Population Association of America*.

Zagheni, E., & Weber, I. (2012). You are where you E-mail: Using E-mail data to estimate international migration rates. In Contractor, N. (Ed.), *WebSci '12: Proceedings of the 4th annual ACM web science conference* (pp. 348–351). New York: Association for Computing Machinery.

Zagheni, E., Weber, I., & Gummadi, K. (2017). Leveraging Facebook's advertising platform to monitor stocks of migrants. *Population and Development Review*, *43*(4), 721-734.