



Ciencias Sociales
Universidad de la República
URUGUAY



Programa de
Población

Universidad de la República
Facultad de Ciencias Sociales
Unidad Multidisciplinaria - Programa de Población

TESIS DE MAESTRÍA EN DEMOGRAFÍA Y ESTUDIOS DE
POBLACIÓN

**Facebook versus encuestas de hogares: oportunidades
y límites de los datos de Facebook para el estudio de la
migración internacional en América Latina**

Camila Montiel

Maestría en Demografía y Estudios de Población

Documento No 25

Febrero 2024

TESIS DE MAESTRÍA EN DEMOGRAFÍA Y ESTUDIOS DE POBLACIÓN

Facebook versus encuestas de hogares: oportunidades y límites de los datos de Facebook para el estudio de la migración internacional en América Latina

Autora: Camila Montiel

Tutora: Dra. Victoria Prieto Rosas (Programa de Población, Universidad de la República)

Co-tutora: Dra. Sofía Gil-Clavel (Delft University of Technology, Países Bajos)

Tribunal:

Dr. Diego Alburez (Max Planck Institute for Demographic Research, Alemania)

Dra. Marcela Cerruti (Centro Nacional de Estudios de Población, Argentina)

Dra. Claudia Masferrer (Centro de Estudios Demográficos, Urbanos y Ambientales - El Colegio de México, México)

Fecha de defensa: 29 de febrero de 2024

Montiel, Camila

Facebook versus encuestas de hogares: oportunidades y límites de los datos de Facebook para el estudio de la migración internacional en América Latina

Tesis Maestría en Demografía y Estudios de Población. - Montevideo: UR.FCS. Unidad Multidisciplinaria. Programa de Población, 2024

78 h, gráfs. cuadros, tablas. (Serie Tesis de Maestría en Demografía y Estudios de Población)

Incluye bibliografía.

Palabras clave: migración internacional, América Latina, Facebook, encuestas de hogares, sesgos.

Resumen

La medición de la migración internacional en América Latina ha enfrentado sistemáticamente un conjunto de restricciones derivadas de la carencia de estadísticas permanentes y armonizadas en tiempo y espacio. De este modo, las Huellas Digitales emergen en el campo de la Demografía como una alternativa complementaria para monitorear en tiempo real los desplazamientos internacionales, pero la evidencia acerca de su validez es aún muy escasa para la región. Los principales antecedentes en la temática, desde la óptica de los métodos de calibración con presencia de datos “reales”, se han planteado como propósito cotejar datos de audiencias de Facebook con estimaciones de stocks migratorios de Naciones Unidas, la plataforma R4V o en un único caso, información censal, y encuentran un nivel elevado de correlación entre las fuentes de información, así como la presencia de sesgos disímiles en términos del país de origen, destino y la composición por edad. A pesar de su creciente uso para el estudio de la migración en la región latinoamericana, las encuestas de hogares no han sido empleadas en este tipo de ejercicios de validación de manera sistemática. El objetivo de este trabajo es analizar en qué medida los datos de usuarios de Facebook etiquetados como “anteriormente viviendo en...” se aproximan a la cuantificación de poblaciones migrantes que arrojan las encuestas de hogares en países latinoamericanos en 2021 y 2022. El análisis multivariado muestra un alto nivel de correlación entre ambos conjuntos de datos, al tiempo que indica diferencias en la magnitud y la dirección de los sesgos según destino y origen. Complementariamente, se evidencia una tendencia general a la sobrerrepresentación de los stocks en Facebook y un mayor ajuste entre fuentes cuando se trata de migrantes absolutos en comparación con los migrantes recientes. A pesar de que los datos de Facebook se ajustan mejor a los de Naciones Unidas que a los de las encuestas, en cualquier caso, se encuentra un buen ajuste respecto a estas fuentes convencionales, lo que refuerza lo encontrado por antecedentes que apuntan a la pertinencia de los datos de Facebook para la estimación de los stocks de migrantes en tiempo real. Este trabajo contribuye a resaltar las oportunidades y limitaciones de los datos de Facebook para la producción de estimaciones de población migrante en tiempo real para el contexto latinoamericano.

Palabras claves: migración internacional, América Latina, Facebook, encuestas de hogares, sesgos.

Abstract

The measurement of international migration in Latin America has systematically faced a set of restrictions derived from the lack of permanent and harmonized statistics in time and space. Digital traces emerge in the field of demography as a complementary alternative to monitor international displacements in real time, but the evidence on their validity is still very scarce for the region. The main antecedents on the subject, from the perspective of calibration methods with the presence of "real" data, have set out to compare Facebook audience data with estimates of United Nations migration stocks, the R4V platform or, in a single case, census information, and have found a high level of correlation between the sources of information, as well as the presence of dissimilar biases in terms of country of origin, destination and composition by age. Despite their growing use for the study of migration in the Latin American region, household surveys have not been used systematically for this type of validation exercises. In this framework, the aim of this work is to analyze the extent to which data from Facebook users labeled as "previously lived in..." approximate the magnitude of the migrant population reported by household surveys in Latin American countries in 2021 and 2022. The multivariate analysis shows a high degree of correlation between the two datasets, while pointing to differences in the magnitude and direction of the biases according to destination and origin. There is also a general tendency to over-represent stocks in Facebook and a better fit between Facebook and household surveys when it comes to absolute migrants compared to recent migrants. Despite the fact that Facebook fits better UN data than surveys data, the results show a good fit respect to traditional data sources, reinforcing previous evidence pointing to the relevance of using Facebook data for the estimation of real time migrant stocks. This work contributes to highlighting the opportunities and limitations of Facebook data for the production of real-time migrant population estimates for the Latin American context.

Key words: international migration, Latin America, Facebook, household surveys, biases.

Agradecimientos

A mis tutoras, Victoria Prieto y Sofía Gil-Clavel, por su enorme predisposición y orientación, el tiempo dedicado a lectura crítica y por el gran aprendizaje que me llevo a partir del trabajo conjunto realizado.

Al apoyo brindado por la Agencia Nacional de Investigación e Innovación (ANII) por medio de la beca de maestría otorgada en el marco del proyecto “Using internet-based data to quantify and sample international migrants. Applications to examine recent immigration to Uruguay” (Programa de Población, Universidad de la República y el Max Planck Institute for Demographic Research, Alemania).

A Martín Pedemonte, por la programación de los extractos de Facebook que representaron el principal insumo para la realización de la tesis.

A Juli, Mari, Gaby y Vic, por el apoyo, sus palabras de ánimo a lo largo del proceso de elaboración de la tesis, y por lo compartido desde 2018, ya que este trabajo también es resultado de lo aprendido a partir del intercambio colectivo diario.

A mi familia y amigas, por su incondicional compañía y apoyo.

Tabla de contenido

| | |
|--|----|
| Introducción | 10 |
| Antecedentes | 13 |
| Revolución del <i>big data</i> y su aplicación al campo de la Demografía | 13 |
| Uso de Huellas Digitales para el estudio de la movilidad internacional | 15 |
| Medición y caracterización de la migración internacional a partir de encuestas de hogares | 19 |
| Marco contextual: el sistema migratorio intrarregional latinoamericano | 22 |
| Objetivos de investigación | 27 |
| Datos y metodología | 28 |
| Descripción de los datos | 28 |
| Datos de Facebook | 28 |
| Encuestas de hogares | 30 |
| Estimaciones de Naciones Unidas | 33 |
| Unidad de análisis | 38 |
| Metodología | 40 |
| Análisis de los sesgos de los datos de Facebook respecto a las encuestas de hogares para migrantes absolutos y recientes | 43 |
| Comparación de los sesgos de los datos de Facebook respecto a las encuestas de hogares y estimaciones de Naciones Unidas | 52 |
| Conclusiones | 60 |
| Limitaciones y futuro trabajo | 63 |
| Referencias bibliográficas | 66 |
| Anexo | 74 |

Índice de cuadros, tablas y gráficos

| | |
|--|----|
| Cuadro 1. Información sobre migrantes absolutos y recientes disponible en las encuestas de hogares por díadas origen-destino según año. Países de destino y origen seleccionados | 32 |
| Cuadro 2. Disponibilidad de cifras de stock de migrantes absolutos según díadas origen-destino en International Migration Stock (Naciones Unidas), 2020 | 34 |
| Cuadro 3. Síntesis de las limitaciones y fortalezas* de los datos de Facebook, encuestas de hogares y estimaciones de Naciones Unidas para la medición de la migración internacional en América Latina | 38 |
| Cuadro 4. Díadas origen-destino que conforman la unidad de análisis para la estimación de los sesgos según definición de migrante. Países de destino y origen seleccionados..... | 39 |
| Cuadro 5. Díadas origen-destino que conforman la unidad de análisis para la estimación de los sesgos de Facebook respecto a las encuestas de hogares y Naciones Unidas. Países de destino y origen seleccionados | 40 |
| Cuadro 6. Especificación de las variables utilizadas en la ecuación 1 y 2 | 42 |
| Cuadro 7. Especificación de las variables utilizadas en las ecuaciones 3 a 5..... | 43 |
| Cuadro 8. Síntesis de la disponibilidad de información sobre migración internacional en las encuestas de hogares de América Latina y el Caribe, 2020-2022 | 74 |
| | |
| Tabla 1. Factor de corrección (proporción de población de 20 a 64 años) de los stocks de migrantes provistos por Naciones Unidas según destino y sexo, 2020 | 35 |
| Tabla 2. Coeficientes del modelo de regresión lineal del logaritmo del stock de migrantes absolutos y recientes de origen i en el destino j con base en las encuestas de hogares. Países seleccionados, 2021 | 50 |
| Tabla 3. Coeficientes del modelo de regresión lineal del logaritmo del stock de migrantes absolutos y recientes de origen i en el destino j con base en las encuestas de hogares. Países seleccionados, 2022 | 51 |
| Tabla 4. Coeficientes del modelo de regresión lineal del logaritmo del stock de migrantes absolutos de origen i en el destino j y sexo k con base en las encuestas de hogares y Naciones Unidas (UN). Países seleccionados, 2020-2022 | 58 |
| | |
| Gráfico 1. Número absoluto de migrantes según fuente de información, definición de migrante empleada y país de origen por año. Países de destino y origen seleccionados, 2020-2022 | 36 |
| Gráfico 2. Relación entre el stock de migrantes del conjunto de datos de Facebook y las estimaciones de migrantes absolutos de encuestas de hogares según país de origen y destino por año. Países seleccionados, 2021 y 2022 | 46 |
| Gráfico 3. Relación entre el stock de migrantes del conjunto de datos de Facebook y las estimaciones de migrantes recientes de encuestas de hogares según país de origen y destino por año. Países seleccionados, 2021 y 2022 | 47 |
| Gráfico 4. Relación entre el stock de migrantes del conjunto de datos de Facebook y las estimaciones de migrantes absolutos de encuestas de hogares según país de origen y destino por sexo. Países seleccionados, 2021 | 54 |
| Gráfico 5. Relación entre el stock de migrantes del conjunto de datos de Facebook y las estimaciones de migrantes absolutos de encuestas de hogares según país de origen y destino por sexo. Países seleccionados, 2022 | 55 |

| | |
|--|----|
| Gráfico 6. Relación entre el stock de migrantes del conjunto de datos de Facebook y las estimaciones de stock de Naciones Unidas (UN) según país de origen y destino por sexo. Países seleccionados, 2020 | 56 |
| Gráfico 7. Porcentaje de personas usuarias de redes sociales según tipo de red social por sexo. Países y redes sociales seleccionadas, 2020..... | 77 |
| Gráfico 8. Porcentaje de personas con conexión a internet en el hogar según país de residencia. Países seleccionados, 2020 | 78 |

Introducción

Una parte constitutiva de la Demografía como campo científico es el proceso de generación de datos a partir de los cuales medir y caracterizar tres fenómenos que afectan al tamaño y composición de la población como lo son la natalidad, mortalidad y migración (Livi-Bacci, 1993; Tapinos, 1988). En este sentido, el estudio de la movilidad internacional ha enfrentado sistemáticamente importantes restricciones para la medición y caracterización del fenómeno debido a la carencia de estadísticas continuas y armonizadas de stocks y flujos migratorios (Cerrutti, Fargues y Awumbila, 2021; Willekens et al., 2016). Asimismo, algunas características intrínsecas de este componente de la dinámica demográfica como su naturaleza repetible, su multidimensionalidad y carácter contextual lo revisten de complejidad (Bengochea, 2018; Willekens et al., 2016).

A pesar de estos desafíos, existe un amplio espectro de fuentes para el estudio de la migración internacional que varían principalmente en función de su propósito, periodicidad de relevamiento y país de consideración, e incluyen registros administrativos de entradas y salidas por puestos fronterizos, censos de población, encuestas de hogares y registros continuos de población (Martínez-Pizarro y Rivera-Orrego, 2016; Calvelo, 2011; Cerrutti, 2009). Específicamente, las últimas tres permiten cuantificar y caracterizar en términos sociodemográficos a los stocks de migrantes, entendidos como la cantidad acumulada de personas nacidas en el extranjero que modificaron su lugar de residencia habitual (Koolhaas, 2023). Los registros continuos de población son de relevancia en el estudio de las migraciones de un grupo selecto de países, donde se tiene un esquema de registro anual de la población a partir del cual se tiene conocimiento sobre los principales acontecimientos vitales de las personas en el territorio (Rampazzo et al., 2021). En cambio, América Latina y el Caribe ha concentrado sus esfuerzos en los censos de población como principal referencia para la construcción de estadísticas de inmigración, emigración y retorno (Masferrer y Prieto, 2019; Bengochea, 2018; Calvelo, 2011; Martínez Pizarro, 2009; Maguid, 2008). Sin embargo, como ha sido señalado en múltiples oportunidades, una de las mayores desventajas que acarrear los censos es la eventual desactualización de la información recolectada dada la periodicidad de su implementación –al menos decenalmente– (Martínez Pizarro, 2009). En el estudio de la movilidad internacional esta desventaja es especialmente notoria, ya que las dinámicas de intercambios poblacionales entre países pueden modificarse sustancialmente en poco tiempo (Bengochea et al., 2021; Martínez Pizarro, 2009). Las encuestas de hogares o de actividad y empleo superan dicha debilidad gracias a su periodicidad anual, pero se enfrentan al desafío de

constituir sus estimaciones a partir de una muestra determinada, y a un número limitado de preguntas destinadas a captar la movilidad internacional (Perdomo Rico, 2022; Gutiérrez et al., 2020; Carrasco y Suárez, 2018; Calvelo, 2011).

En este escenario caracterizado por los desafíos que revisten las fuentes antes mencionadas para captar la migración de manera fiable y temporalmente oportuna, han surgido recientemente una serie de investigaciones que analizan la intensidad de los flujos y las características de los stocks de poblaciones en situación de movilidad aprovechando la información que deja el uso de dispositivo móviles y del internet. El creciente campo de la Demografía Digital ha aprovechado la información generada a través del uso de plataformas de redes sociales web, dispositivos móviles o buscadores de internet, no sin advertir sobre los sesgos propios de esta información (Bengochea, 2022; Kashyap et al., 2022; Kashyap, 2021; Alburez et al., 2019; Billari y Zagheni, 2017). Entonces, en un contexto de restricciones metodológicas como las señaladas anteriormente, es pertinente valorar en qué medida el aprovechamiento de las Huellas Digitales son una ventana de oportunidad para refinar las estadísticas de migración internacional en la región latinoamericana en tanto proporcionan insumos en tiempo real, de manera rápida y a bajo costo (Rampazzo et al., 2021; Alburez et al., 2019).

En América Latina, algunos antecedentes han realizado el ejercicio parcial de contrastar las Huellas Digitales con información derivada de estimaciones de Naciones Unidas (Prieto Rosas et al., 2022b; Spyrtatos et al., 2019), censos de población (Varona, 2022) o registros administrativos (Palotti et al., 2020). Y más allá del ejercicio realizado por Rosati, Cerrutti y Maguire (2021) para el caso de nacionales venezolanos en la Ciudad Autónoma de Buenos Aires con base en la Encuesta Anual de Hogares, de momento no se encontraron antecedentes de validación de datos de redes sociales web que se apoyen en las encuestas de hogares y abarquen un amplio conjunto de orígenes y destinos. Esto marca una importante diferencia con la producción académica generada desde países del norte global, donde las encuestas de hogares han sido tomadas como principal referencia para cotejar la información sobre migrantes extraída de redes sociales web (Rampazzo et al., 2021; Spyrtatos et al., 2019; Zagheni, Weber y Gummadi, 2017). Por otro lado, la validación de las Huellas Digitales apremia en el escenario latinoamericano motivada por la necesidad de recolectar información en tiempo real sobre los desplazamientos de población en la región, donde se corrobora una mayor participación de personas con necesidad de protección internacional y/o en tránsito que son captadas débilmente por censos o encuestas multipropósitos (OIM, 2021, 2022a; Herrera y Gómez, 2022).

En este marco, la pregunta de investigación que guía el presente trabajo es en qué medida los datos de usuarios etiquetados como “anteriormente viviendo en...”¹ de Facebook se aproximan a la cuantificación de poblaciones migrantes que arrojan las encuestas de hogares en países latinoamericanos en 2021 y 2022. En particular, el análisis buscó responder a una serie de preguntas específicas referidas al ejercicio de validación: ¿Cuál es la dirección y magnitud de los sesgos entre los datos de Facebook y las encuestas de hogares según país de destino y origen? ¿En qué medida dichos sesgos se aproximan a los estimados para los datos de Facebook y Naciones Unidas? ¿Qué definición de migrante se asemeja más a la cuantificación de poblaciones migrantes que provee Facebook? ¿En qué medida los sesgos entre fuentes varían según sexo?

Con este propósito, se construyeron dos bases de datos que reúnen estimaciones armonizadas de stocks de migrantes a partir de microdatos de encuestas de hogares de nueve países de destino de la región, cifras de usuarios diarios de Facebook y estimaciones de *International Migration Stock* (Naciones Unidas, 2020), según i) país de origen y destino y ii) país de origen, destino y sexo. Con base en la información recolectada, se ajustaron una serie de modelos lineales multivariados para estimar la correlación y sesgos entre fuentes. El presente trabajo ofrece un ejercicio de análisis sistemático de los sesgos de Facebook contra dos fuentes de naturaleza más convencional –encuestas de hogares y Naciones Unidas– que busca contribuir al acervo sobre validación de Huellas Digitales para la medición en tiempo real de la migración intrarregional en varios países de América Latina.

El documento se estructura en cinco secciones. Primero, se exponen una serie de antecedentes de relevancia en el marco de la Demografía Digital, con especial foco en los ejercicios de contraste de Huellas Digitales y fuentes de información tradicionales para la cuantificación de la migración internacional. Segundo, se presenta una síntesis de las principales características de la movilidad internacional intrarregional en América Latina en el período reciente. Tercero, se detallan los objetivos de la investigación, así como los datos y métodos empleados en su concreción. Cuarto, se desarrollan los principales resultados del

¹ Variable provista por Facebook Ads Manager que vincula el lugar de residencia actual (país de destino) y anterior (aquél mencionado en la etiqueta que lleva por nombre “Han vivido en *j* (anteriormente expatriados - *j*)”) del usuario: <https://developers.facebook.com/docs/marketing-api/audiences/reference/basic-targeting#behaviors> (Acceso el 26 de enero de 2024). Con el propósito de simplificar la lectura del presente texto, y siguiendo el término que emplea la mayor parte de la literatura, se hará un uso indistinto de *expats* y “anteriormente viviendo en...” para hacer alusión a la clasificación provista por Facebook mencionada anteriormente.

estudio en clave del análisis de los sesgos entre Facebook y las encuestas de hogares según definición de migrante y la comparación de ellos respecto a Facebook y Naciones Unidas para migrantes absolutos. También en esta sección se discute la significancia de las diferencias por sexo. Por último, se esbozan una serie de conclusiones y reflexiones finales.

Antecedentes

La presente sección tiene como objetivo presentar la evidencia acumulada sobre ejercicios de validación de Huellas Digitales, con especial foco en aquellas derivadas de Facebook, los cuales han realizado un análisis exhaustivo de la dirección y magnitud de los sesgos con base en distintas fuentes de información tradicionales. Como se desarrollará inmediatamente, la literatura que ha analizado contextos de destino latinoamericanos es aún incipiente, y solo en dos oportunidades se han utilizado como estándar de comparación a los censos de población (Varona, 2022) y las encuestas de hogares (Rosati, Cerrutti y Maguire, 2021). Por ende, también es de relevancia conocer las principales fortalezas y debilidades que se ha documentado presentan las encuestas de hogares de América Latina considerando que será la fuente de información convencional empleada como estándar de comparación en el presente análisis. De este modo, la sección se organiza en tres grandes ejes que buscan i) presentar las principales características de las Huellas Digitales en tanto fuente de información de relevancia para la investigación demográfica y las áreas de estudio en las que se han empleado; ii) los principales antecedentes directos de cotejo de datos digitales y tradicionales para la medición de la migración internacional; y iii) una sucinta descripción del contenido y desafíos de las encuestas de hogares en la región para el estudio de la movilidad internacional.

Revolución del *big data* y su aplicación al campo de la Demografía

La emergencia del *big data* como fuente de información complementaria para la investigación demográfica es muy reciente, y se encuentra impulsada por transformaciones a nivel tecnológico y de conectividad, pero su definición no está exenta de discusiones y existen límites difusos sobre qué datos pueden ser catalogados como *big data* y cuales no (Cesare et al., 2018; Salganik, 2018; Zagheni y Weber, 2015). No obstante, parece identificarse cierto consenso sobre que el *big data* se constituye a partir de la conjugación de tres elementos distintivos que son el volumen de la información almacenada, la variedad de estructuras de los datos y la gran velocidad con la que se producen, los cuales se han denominado las “tres Vs” (Cesare et al., 2018). Es preciso destacar que lo novedoso es la simultaneidad de los tres aspectos dado que

el volumen de información es un elemento relativamente constante en el campo de los estudios demográficos (Kashyap, 2021; Billari y Zagheni, 2017) y, en particular, es el componente de la velocidad de producción de la información el que demarca una importante diferencia con las fuentes de datos tradicionales (Kashyap, 2021). Asimismo, una de las principales ventajas del uso de este tipo de información radica en que su producción, extracción y almacenamiento puede darse en tiempo real y de este modo es posible monitorear cambios de ciertos comportamientos demográficos en momentos precisos utilizando Huellas Digitales, entendidas como los datos generados a partir de la interacción de las personas con los medios digitales, como pueden ser las redes sociales web.

Lo anterior supone una notoria diferencia con las fuentes de índole tradicional como los censos de población y encuestas de hogares para las cuales la publicación de los resultados se da de manera más dilatada y por períodos más prolongados (Cesare et al., 2018). Más allá de estas características que se pueden definir como positivas, Salganik (2018) y Kashayap (2021) enumeran una serie de desafíos que trae aparejado el uso de Huellas Digitales con fines de investigación, entre los que se destacan i) el ser incompletos, ya que el hecho de que su producción no esté guiado por preguntas de investigación concretas puede generar que no se tenga disponible la información requerida; iii) el ser no representativos, en la medida de que no se pueden realizar generalizaciones a la población general; iv) la falta de poder de control del investigador ante cambios en los instrumentos de medición o en los algoritmos de construcción de variables; y v) la presentación poco organizada que demanda un proceso de limpieza y estructuración de la información extraída.

En este escenario, la Demografía gracias a su particular preocupación por la calidad de los datos y las fuentes de información, ha contribuido notoriamente al estudio de los límites y ventajas del uso de las Huellas Digitales con fines de investigación en las Ciencias Sociales. Las principales contribuciones en este sentido pueden organizarse en dos grupos; por un lado, las derivadas de los trabajos que se interesaron por comprender los sesgos de índole demográfica en el uso de las redes sociales web y, por otro lado, aquellas donde se ha puesto el foco en el estudio de aspectos relativos a la dinámica demográfica.

En primer lugar, uno de los principales desafíos que enfrenta el uso de los Huellas Digitales refiere a la existencia de sesgos de selección, que resultan del acceso y uso estratificado de las redes sociales web o dispositivos móviles según características sociodemográficas (Hargittai, 2018; Pestre, Letouzé, y Zagheni, 2016). Específicamente, la literatura reciente identifica un uso más intensivo de redes sociales entre individuos con nivel educativo e ingresos más altos, personas blancas, mujeres y jóvenes (Donzowa, Perrotta y

Zagheni, 2023; Gil-Clavel, Zagheni y Bordone, 2022; Gil-Clavel y Zagheni, 2019; Hargittai, 2015, 2007). En particular, los usuarios de Facebook tienen asociada una mayor probabilidad de ser de etnia blanca o de tener un mayor nivel educativo y de ingresos con respecto a los usuarios de Google (Stern et al 2017).

En segundo lugar, el uso de Huellas Digitales en el marco de los estudios demográficos comienza a tener lugar con posterioridad a 2010, y son variados los temas de investigación en donde es posible identificar su aplicación, tales como la caracterización de patrones de fecundidad (Wilde et al., 2020; Rampazzo et al., 2018; Ojala et al. 2017; Billari, D'amuri, y Marcucci, 2016; Markey y Markey 2013), las preferencias en las intenciones de conformación de pareja (Hitsch, Hortaçsu, y Ariely, 2010) y los aspectos relativos a la mortalidad o epidemiología (Basellini et al., 2021; Grow et al., 2020). Una de las áreas de investigación que cuenta con mayor caudal de evidencia acumulada refiere al estudio de la migración. En ella, las Huellas Digitales se han usado tanto para estudiar cómo se integran las poblaciones migrantes a los países de acogida (Gil-Clavel, Grow y Bijlsma, 2023; Vieira et al., 2020; Stewart et al., 2019), así como los patrones de movilidad internacional (Varona, 2022; Prieto Rosas et al., 2022b; Rampazzo et al., 2021; Alexander, Polimis y Zagheni, 2020; Hsiao et al., 2020; Palotti et al., 2020; Spyratos et al., 2019, 2020; Zagheni et al., 2018; Zagheni, Weber y Gummadi, 2017), la cual es de interés particular para el presente trabajo. En la siguiente sección se detallan los principales hallazgos encontrados en referencia al uso de Huellas Digitales para la medición de la migración internacional.

Uso de Huellas Digitales para el estudio de la movilidad internacional

En lo concerniente al componente de la dinámica demográfica que es de interés para el presente trabajo, el uso de Huellas Digitales es relativamente novedoso y ha sido más intensivo en contextos de acogida del norte global (Rampazzo et al., 2021; Wanner, 2021; Alexander, Polimis y Zagheni, 2020; Hsiao et al., 2020; Wisniowski et al., 2020; Del Fava, Wiśniowski, y Zagheni, 2019; Spyratos et al., 2018, 2019; Zagheni et al., 2018; Gendronneau et al., 2019; Fiorio et al., 2017; Zagheni, Weber y Gummadi, 2017; Hughes et al., 2016). En cambio, en América Latina y el Caribe la producción científica sobre este tema es más reciente y menos abundante (Varona, 2022; Prieto Rosas et al., 2022a, b; Rosati, Cerrutti y Maguire, 2021; Palotti et al., 2020; Spyratos et al., 2019). En cualquiera de estos contextos, una proporción no menor de los trabajos académicos se han concentrado en el estudio de los stocks migratorios a partir de las cifras de audiencias que Facebook Ads Manager reporta como “Han vivido en *j*

(anteriormente expatriados - *j*)” (Varona, 2022; Prieto Rosas et al., 2022b; Rampazzo et al., 2021; Rosati, Cerrutti y Maguire, 2021; Alexander, Polimis y Zagheni, 2020; Hsiao et al., 2020; Palotti et al., 2020; Spyratos et al., 2019, 2020; Zagheni et al., 2018; Zagheni, Weber y Gummati, 2017) y entre ellos es posible identificar dos grandes grupos que se diferencian en su principal cometido. El primero, integrado por aquellos trabajos que buscan contrastar la información digital con datos provenientes de Naciones Unidas, censos de población, encuestas de hogares o registros administrativos (Varona, 2022; Prieto Rosas et al., 2022b; Rampazzo et al., 2021; Rosati, Cerrutti y Maguire, 2021; Spyratos et al., 2019, 2020; Palotti et al., 2020; Zagheni, Weber y Gummati, 2017). El segundo grupo, compuesto por los que se han concentrado en combinar los datos proporcionados por Facebook Ads Manager con información de encuestas de hogares como la *American Community Survey* o la *Labour Force Survey* por medio de modelos bayesianos para alcanzar estimaciones y proyecciones ajustadas de stocks (Alexander, Polimis y Zagheni, 2020; Rampazzo et al., 2021).

A la hora de trabajar con datos digitales es indispensable lidiar con los sesgos de selección, y una alternativa para su tratamiento es la comparación con fuentes de información de tipo tradicional que sirvan de insumo para validar los resultados alcanzados (Bengochea, 2022). Específicamente, como fue mencionado en la sección precedente, en el campo de estudio de la migración internacional las fuentes de información disponibles acarrearán una serie de desventajas que deben ser tenidas en cuenta a la hora de estimar stocks y flujos migratorios, lo cual implica que incluso aquellos parámetros que se suelen tomar como comparación no dejan de estar afectados por efectos de inclusión y exclusión. Más allá de esta salvedad, como menciona Bengochea (2022), el trabajo de Zagheni y Weber (2012) representa la primera contribución a la literatura sobre la estimación de los stocks migratorios mediante el tratamiento de los sesgos presentes en los datos digitales, pero es el artículo de Zagheni, Weber y Gummati (2017) el que demarca el inicio del uso de datos de Facebook para el estudio de la movilidad internacional.

Zagheni, Weber y Gummati (2017) comparan las cifras de usuarios de Facebook etiquetados como *expats* residiendo en Estados Unidos con información derivada de la *American Community Survey*, y mediante una regresión lineal multivariada que busca explicar el logaritmo del número de inmigrantes por sexo, edad y origen resultante de la segunda fuente mencionada, encuentran un nivel de asociación importante entre ambos tipos de datos, que alcanza un valor del R-cuadrado ajustado de 69,7%. No obstante, los autores identifican que los datos de Facebook sobreestiman el stock de los grupos etarios más jóvenes mientras que lo inverso sucede con las edades más avanzadas, y subestiman la magnitud del stock para la

mayoría de los orígenes nacionales. Complementariamente a la contribución en términos de la validación de los datos digitales, el trabajo de Zagheni, Weber y Gummadi (2017) representa un antecedente de referencia para la metodología de modelización de los sesgos, que ha sido tomada como referencia por otros autores como Spyrtos et al. (2019), Prieto Rosas et al. (2022b) y Varona (2022).

En particular, Spyrtos et al. (2019) encuentran un alto grado de ajuste entre los datos de usuarios *expats* de 15 a 64 años extraídos de Facebook y las cifras de migrantes arrojadas por las estimaciones de población de Naciones Unidas, los datos de *Eurostat* y la *American Community Survey* según sexo, país de residencia actual y anterior. A diferencia de Zagheni, Weber y Gummadi (2017), los autores realizan un paso previo a la estimación del modelo de regresión que refiere a la corrección del número de *expats* por medio de la aplicación de una tasa de penetración de Facebook en el país de residencia actual y previo del individuo. Cabe señalar, que Spyrtos et al. (2019) encuentran diferencias en los niveles de asociación entre las fuentes según origen nacional, como es el caso de los latinoamericanos en Estados Unidos, quienes se encuentran sobrerrepresentados en Facebook respecto a la *American Community Survey*.

Por su parte, Prieto Rosas et al. (2022b) trabajan con información de extractos diarios de usuarios anteriormente viviendo en 13 orígenes que reportan residir en 21 países de destino latinoamericanos al momento de la extracción, comprándolos con datos de stocks de migrantes de Naciones Unidas para 2019 y 2020. Las autoras especifican un modelo de regresión lineal cuya variable dependiente es el logaritmo de la proporción de inmigrantes en la población total según Naciones Unidas. Entre los principales resultados de este trabajo, se destaca que los datos digitales tienen un nivel de asociación de entre 83% y 92% con las cifras de Naciones Unidas, y que existen diferencias significativas en las proporciones reportadas por ambas fuentes según año de referencia, origen nacional y tramos etarios. En particular, los modelos estimados por Prieto Rosas et al. (2022b) muestran que Facebook tiende a sobrerrepresentar los stocks de migrantes respecto a Naciones Unidas para la amplia mayoría de los países de destino considerados en el análisis, mientras que los sesgos según país de origen parecen mostrar la dirección contraria. Además, de acuerdo a estas estimaciones, las diferencias entre ambas fuentes no son significativas según sexo (Prieto Rosas et al., 2022b). Por otro lado, una de las principales contribuciones de Prieto Rosas et al. (2022b), que coincide con la sugerencia planteada por Gil-Clavel, Zagheni y Bordone (2022), está dada por la utilización del número de usuarios activos diarios (DAU por sus siglas en inglés) de Facebook en lugar del número de

usuarios activos mensuales (MAU por sus siglas en inglés)², lo cual es de relevancia cuando se trabaja con comunidades de origen pequeñas que quedan invisibilizadas en el MAU producto de que éste solo arroja cifras para orígenes con un número de usuarios superior a mil.

Palotti et al. (2020) también encuentran un nivel de asociación elevado entre los datos digitales y tradicionales al considerar el número de usuarios activos mensuales de Facebook cuyo lugar de residencia anterior era Venezuela y las estimaciones de migrantes de dicho origen proporcionadas por la Plataforma de Coordinación Interagencial para Refugiados y Migrantes de Venezuela (R4V). Esto cobra especial pertinencia en un escenario donde las estimaciones de stocks de migrantes provistas por Naciones Unidas comienzan a incluir datos sobre desplazados venezolanos para ciertos países de América Latina a partir de 2020.

En otro orden, es de relevancia destacar el trabajo de Varona (2022) por dos motivos; es el primero en utilizar información censal como estándar de comparación y plantea la discusión sobre la definición de migrante que más se aproxima a la etiqueta *expats*. Con base en extracciones de datos de usuarios de Facebook en México, evalúa el grado de ajuste de la categoría *expats* con seis definiciones de migración resultantes de la combinación de las variables lugar de nacimiento, lugar de residencia cinco años antes y temporalidad del desplazamiento que obtiene del Censo de Población y Vivienda de 2020 de México. La autora encuentra que en general existe un elevado grado de ajuste entre los datos de Facebook y el censo, que alcanza su valor máximo para la definición de migrantes absolutos según país de nacimiento, con un valor del R^2 ajustado de 80,8%. Con relación a la migración reciente, la correlación es mayor cuando se clasifica a las personas migrantes según país de residencia cinco años antes respecto a la definición según país de nacimiento (Varona, 2022). Por otro lado, los modelos generales estimados por Varona (2022) muestran una tendencia general a la subrepresentación de los datos de Facebook respecto al censo, aunque para ciertos orígenes, como República Dominicana, la relación es inversa. Además, la autora no encuentra diferencias significativas entre fuentes según sexo para ninguna de las tres definiciones mencionadas, una vez que se controla por la edad, el país de nacimiento o el de residencia cinco años antes, y el logaritmo de la proporción de *expats*.

Por último, un antecedente directo del presente trabajo es el ejercicio de validación de datos de Facebook realizado por Rosati, Cerrutti y Maguire (2021) para nacionales venezolanos en la Ciudad Autónoma de Buenos Aires. Los autores comparan las cifras de usuarios

² <https://developers.facebook.com/docs/marketing-api/reference/application/> (Acceso 26 de enero de 2024).

etiquetados como anteriormente viviendo en Venezuela con el stock de población de dicho origen en la Encuesta Anual de Hogares, y encuentran que en términos generales Facebook sobrerrepresenta la magnitud de dicho contingente poblacional, así como presenta brechas en la estimación según tramos etarios. En particular, los datos digitales subrepresentan a las personas de 13 a 24 años, mientras que lo inverso se identifica para aquellos de 25 y más años, siendo más amplia la brecha para el grupo de 25 a 44 años (Rosati, Cerrutti y Maguire, 2021).

Con base en la revisión de antecedentes expuesta es posible afirmar que la medición y caracterización de la migración internacional en América Latina a partir de Huellas Digitales es un área de estudio de incipiente desarrollo, donde las principales publicaciones de referencia han definido como propósito validar los datos de la Facebook Ads Manager con información censal, estimaciones de Naciones Unidas o de la plataforma R4V o en casos puntuales respecto a encuestas de hogares (Varona, 2022; Prieto Rosas et al., 2022b; Rosati, Cerrutti y Maguire, 2021; Palotti et al., 2020; Spyrtatos et al., 2019). Los ejercicios de validación para países de origen y/o destino latinoamericanos muestran i) un elevado ajuste entre las Huellas Digitales y las fuentes de tipo convencional empleadas como estándar de comparación, ii) una tendencia general a la subrepresentación de los stocks de migrantes según país de origen aunque con excepciones cuando se toma como referencia datos de encuestas de hogares; y iii) diferencias no significativas entre fuentes según sexo. Por otro lado, la validación sistemática de los datos digitales contra las encuestas de hogares es una tarea pendiente en América Latina, y representa una contribución de interés por al menos tres motivos: i) es un ejercicio que ha mostrado resultados positivos para otras regiones; ii) para buena parte de los países latinoamericanos se cuenta con actualizaciones anuales de las encuestas de hogares o población activa; iii) las encuestas permiten lidiar con algunos de los sesgos presentes en otras fuentes de información como la exclusión de personas migrantes en situación documental irregular en los registros administrativos (i.e. cifras de stock de personas migrantes de origen venezolano provistas por la plataforma R4V) y las estimaciones demográficas bajo supuestos de evolución del saldo migratorio que generalmente toman el valor nulo (i.e. *International Migrant Stock* de Naciones Unidas).

Medición y caracterización de la migración internacional a partir de encuestas de hogares

Las encuestas de hogares se nutren de los censos de población para la construcción de sus marcos muestrales y factores de expansión, y por lo general comparten con ellos un conjunto de preguntas semejantes que permiten caracterizar las estructuras de las poblaciones según

variables demográficas claves como el sexo, edad, país de nacimiento, lugar de residencia, entre otras. Asimismo, uno de los principales aspectos a destacar de este tipo de encuestas radica en que generalmente cuentan con una gama muy amplia de preguntas vinculadas a las condiciones de vida de las poblaciones, lo que habilita a caracterizar de manera detallada el acceso al empleo, la vivienda, los servicios de educación, salud y el esquema de protección social de cada país (Carrasco y Suárez, 2018). En este escenario, buena parte de los trabajos precedentes que han utilizado a las encuestas de hogares como fuente de información para el estudio de la migración internacional, han puesto foco en el análisis comparado de los desempeños en diferentes dimensiones del proceso de inclusión socioeconómica entre la población nativa e inmigrante (Prieto Rosas y Montiel, 2020; Prieto Rosas y Márquez Scotti, 2019; Carrasco y Suárez, 2018; Morales Gamboa, 2008). Sin embargo, es pertinente destacar los aportes de dos antecedentes recientes que se han planteado como propósito analizar la disponibilidad y calidad de la información de las encuestas de hogares para el estudio de la migración internacional en América Latina, de manera tal que contribuyen específicamente a la discusión sobre la relevancia de esta fuente para la medición de la movilidad internacional en la región. Asimismo, el uso de las encuestas de hogares como fuente para la medición de la migración internacional en América Latina es un muy reciente, y junto con las encuestas específicas ganó especial relevancia en el actual período intercensal ante la carencia de otras fuentes a partir de las cuales monitorear las tendencias de la movilidad internacional en la región (Prieto Rosas y Zapata, 2023; OIM, 2023³; Giorguli et al., 2023; Carrasco y Baizán, 2023; Prieto Rosas et al., 2022; Perdomo Rico, 2022; Márquez, Prieto Rosas y Escoto, 2020; Gutiérrez et al., 2020; Carrasco y Suárez, 2018). Por otro lado, cabe recordar que una de las principales virtudes de las encuestas de hogares radica en su periodización de recolección anual (Perdomo Rico, 2022; Gutiérrez et al., 2020), lo cual pondera su utilidad para monitorear las modificaciones en las tendencias de los flujos migratorios en períodos de corta duración.

En primer lugar, Gutiérrez et al. (2020) disponen su análisis para 18 encuestas de hogares de países de la región en 2017, de las cuales 15 cuentan con algún tipo de pregunta comparable para identificar a las personas migrantes. Considerando este segundo conjunto de casos nacionales y en función de seis requisitos estadísticos que definen el umbral mínimo de precisión de las estimaciones, los autores encuentran que para la totalidad de países el

³ En los últimos años se realizaron múltiples *Displacement Tracking Matrix* (DTM) en diez países de residencia sudamericanos, con foco en captar los flujos vulnerables de personas migrantes de origen venezolano y cubano (OIM, 2023).

porcentaje de personas inmigrantes en el total de la población alcanza estimaciones precisas excepto para Bolivia, Honduras y Nicaragua. Sin embargo, cuando analizan el nivel de precisión de un conjunto de indicadores sociodemográficos según condición migratoria, encuentran que las restricciones son mayores y sólo es posible tener estimaciones precisas para alguno de ellos en el caso de Argentina, Brasil, Chile, Colombia, Costa Rica, Guatemala, Panamá, y República Dominicana. Complementariamente, Gutiérrez et al. (2020) definen una serie de desafíos asociados al uso de las encuestas de hogares para la medición y caracterización de las poblaciones migrantes en la región: i) presencia de tamaños muestrales insuficientes para contar con estimaciones precisas según varios niveles de desagregación en simultáneo; ii) eventual desactualización de los marcos muestrales cuando suceden importantes modificaciones en la magnitud y composición de los flujos migratorios en el período intercensal; iii) subestimación de los stocks por la no consideración de las viviendas colectivas en los diseños muestrales⁴ o por el “ocultamiento” de ser migrante por parte de las personas encuestadas⁵.

En segundo lugar, Perdomo Rico (2022) encuentra resultados consistentes con los de Gutiérrez et al. (2020) en relación con la disponibilidad de preguntas que habilitan la identificación de las personas migrantes en las encuestas de hogares de América Latina para los años posteriores a 2017. A partir de la sistematización del contenido de las encuestas, se reconocen tres formas de identificación de las personas migrantes, que refieren al lugar de nacimiento, la nacionalidad y el lugar de residencia de la madre al momento del nacimiento del/la encuestado/a, siendo la primera la más adoptada entre los países latinoamericanos y caribeños (Perdomo Rico, 2022). Por otra parte, el autor evalúa la calidad de las estimaciones de cada encuesta a través de dos ejercicios metodológicos. Primero, tomando como referencia la metodología implementada por Gutiérrez et al. (2020), analiza la calidad de los datos por medio del nivel de representatividad de la población migrante en la muestra y sostiene que

⁴ Se ha documentado que las viviendas colectivas representan una alternativa habitacional de relevancia para personas migrantes de ciertos orígenes latinoamericanos en ciudades de la región, como es el caso de las personas haitianas en Santiago (Edwards y Greene, 2022) o dominicanas y cubanas en Montevideo (Bengochea y Madeiro, 2020).

⁵ Las personas migrantes suelen ser incluidas dentro de las poblaciones ocultas o de difícil acceso, que se caracterizan por la dificultad para obtener muestras representativas por motivo de la carencia de un marco muestral de referencia que las contemple o bien por la reticencia a la participación en investigaciones producto de procesos de estigmatización (i.e. personas migrantes en situación documental irregular) (Cárdenas Castro y Yáñez Yáñez, 2012). En este contexto, vale la pena destacar dos ejercicios metodológicos que intentan superar los obstáculos descritos, como lo son el Respondent-Driven Sampling (RDS) (Tyldum y Johnston, 2014) y las encuestas en línea que utilizan a Facebook como espacio de promoción y reclutamiento (Donzowa, Perrotta y Zagheni, 2023).

Colombia, Costa Rica, Chile, Ecuador, Honduras, México, Panamá y República Dominicana ofrecen estimaciones de alta calidad, mientras que Argentina, Bolivia, Guyana, Paraguay, Perú, Venezuela alcanzan el nivel medio. Segundo, coteja el valor de la participación relativa de la población migrante en la población total de las encuestas de hogares respecto a la estimación provista por el Departamento de Asuntos Económicos y Sociales de las Naciones Unidas, y muestra que en promedio la magnitud de la primera fuente de información es inferior a su patrón de comparación.

Para concluir, se puede afirmar que las encuestas de hogares son una fuente de información relevante para conocer la magnitud y participación relativa de la población migrante en los períodos intercensales producto de su levantamiento anual, pero enfrentan importantes restricciones derivadas de su tamaño muestral cuando se busca hacer análisis pormenorizados para más de una variable, la subestimación de los contingentes poblacionales por omisión de viviendas colectivas o la presencia de grupos de “difícil acceso” y la desactualización de los marcos muestrales (Perdomo Rico, 2022; Gutiérrez et al., 2020). Cabe señalar que una de las estrategias implementadas para intentar superar este obstáculo refiere al análisis simultáneo de muestras para varios años consecutivos (*pooled data*) (Preston, Heuveline y Guillot, 2001). Consecuentemente, es necesario advertir que, incluso el estándar de comparación a partir del cual se validarán las cifras de audiencias de Facebook en esta oportunidad no está exento de limitaciones para estudiar la migración internacional en la región.

Marco contextual: el sistema migratorio intrarregional latinoamericano

La presente sección tiene como propósito describir las principales características de la migración intrarregional en América Latina en el período reciente, de modo tal que no es objeto de este apartado realizar una revisión histórica de los intercambios poblacionales al interior de la región.

La migración intrarregional en América Latina presenta un contexto heterogéneo si se tiene en cuenta la direccionalidad de los flujos migratorios, los cuales, en función de su intensidad y permanencia temporal, dan lugar a tres subsistemas migratorios delimitados por i) países de Centroamérica y México, ii) países del Caribe, y iii) aquellos situados en América del Sur (Bengochea, 2018). En esta oportunidad se describirán únicamente los dos últimos subsistemas ya que son los que aluden a los países de destino que son objeto de estudio en el presente análisis (ver sección “Datos y metodología”).

En primer lugar, los intercambios poblacionales entre México y países próximos de Centroamérica como Costa Rica, El Salvador, Guatemala, Honduras, Nicaragua y Panamá están motivados fundamentalmente por factores laborales y con una marcada dirección sur-norte, donde Estados Unidos se posiciona como el destino final deseado por buena parte de las personas en situación de movilidad (Bengochea, 2018; Nájera, 2016). En este contexto, México se convierte en un espacio de tránsito o bien de permanencia temporal clave de migrantes de origen guatemalteco, hondureño o salvadoreño, que encuentran en la frontera de dicho país con Guatemala un espacio estratégico para llegar a Estados Unidos, muchas veces bajo situaciones irregulares (Nájera, 2016). En particular, se destacan dos flujos migratorios de relevancia en dicho subsistema: el de personas nicaragüenses a Costa Rica y el de migrantes de origen guatemalteco a México (Canales Cerón y Rojas Weisner, 2018). En 2020 siete de cada diez inmigrantes en Costa Rica eran de origen nicaragüense (Naciones Unidas, 2020a), y se ha documentado que se trata de una migración impulsada por factores laborales, concentrada en edades activas, con una creciente feminización de los flujos y que se inserta en nichos laborales del sector agrícola y el trabajo doméstico (Castro Valverde, 2007; Gatica López, 2007; Baumeister, 2006). Por su parte, las personas migrantes nacidas en Guatemala representaban el 5% de la población extranjera que residía en México en 2020, siendo el segundo origen de relevancia luego de Estados Unidos (64%) (Varona, 2022). Al respecto, existe una doble lógica de movilidad: aquella que refiere a la migración transfronteriza de trabajadores que responde fundamentalmente a la demanda de mano de obra en el sector agrícola del estado de Chiapas (Nájera, 2017) y, por otro lado, el creciente tránsito irregular que tiene como destino final Estados Unidos, y donde los desplazamientos asumen generalmente la forma de “caravanas” (Gandini, Fernández de la Reguera y Narváez Gutiérrez, 2020). Es importante destacar que, para la última ronda censal de México, emerge un tercer origen de importancia que es Venezuela y que alcanza una participación relativa semejante a la de Guatemala (Varona, 2022).

En segundo lugar, se encuentra el subsistema migratorio de América del Sur, el cual se vio dinamizado por la imposición de barreras a migrantes en situación documental irregular para el ingreso a países del norte global, la disminución de oportunidades laborales en contextos de destinos tradicionales, la mejora en la posición económica relativa de países de la región, y los avances en materia de política migratoria en los países sudamericanos (Bengochea, 2018; Cerruti y Parrado, 2015). Al momento de la ronda censal de 2010 el escenario migratorio sudamericano tenía como epicentros históricos de la movilidad intrarregional a Argentina y Venezuela, y a Chile y Ecuador como destinos emergentes (Bengochea, 2018; Cerruti y

Parrado, 2015; Martínez Pizarro, Cano Christiny y Soffia Contrucci, 2014) pero como se verá inmediatamente, esta configuración comienza a cambiar radicalmente a partir de 2015 con la ocurrencia del éxodo venezolano. Hasta entonces, la composición de los flujos por orígenes nacionales se caracterizaba por una fuerte preeminencia de los intercambios fronterizos (Bengochea, 2018). En el caso de Argentina, uno de los orígenes más numerosos y con mayor presencia histórica ha sido el de nacionales paraguayos, siendo especialmente notorio el incremento de las personas migrantes nacidas en Perú y Bolivia en los años próximos a la ronda censal 2010 (Cerruti, 2009; Pellegrino, 2009). Por otro lado, la composición por orígenes en Chile ha estado liderada por nacionales argentinos, peruanos y bolivianos (Stefoni, 2011; Martínez Pizarro, 2005), mientras que, en Colombia se ha destacado la afluencia de personas de origen venezolano y la movilidad bilateral con Ecuador (Bengochea, 2018). La inmigración con destino a Perú ha procedido principalmente de Chile, Colombia y Argentina (Sánchez Aguilar, 2012), mientras que Uruguay ha tenido una preeminencia histórica reciente de recepción de migrantes de sus países fronterizos –Argentina y Brasil– (Bengochea, 2014).

Un aspecto sobresaliente de la movilidad internacional en América Latina y el Caribe en la última década refiere al aumento de los desplazamientos entre países de la región, en detrimento, fundamentalmente, de los flujos hacia Estados Unidos (PNUD, 2023; Vera y Adler, 2020). Específicamente, se ha observado que el stock de migrantes intrarregionales poco más que se duplicó entre 2010 y 2020, siendo especialmente elevado el incremento en el último quinquenio (6.2 y 11.3 millones de migrantes en 2015 y 2020 respectivamente) (PNUD, 2023). Estas tendencias al alza se explican fundamentalmente por dos factores: el inicio del éxodo venezolano en 2015, que ha tenido como principales destinos a países de la región, y el endurecimiento de la política migratoria estadounidense (Bengochea y Pellegrino, 2023; PNUD, 2023). Venezuela pasó de ser uno de los principales epicentros de la movilidad intrarregional sudamericana y por ende de tener un perfil receptor a ser netamente expulsor – se estima que alrededor de 7.71 millones de nacionales venezolanos se encuentran fuera de su país de origen en 2023 (Plataforma R4V, 2023). En cambio, países como Colombia, Ecuador y Perú pasaron a tener tasas netas de migración positivas por primera vez desde mediados del siglo XX (Prieto Rosas y Zapata, 2023). De este modo, Venezuela se ha convertido en el principal país de origen que compone la población inmigrada para buena parte de los contextos de destino sudamericanos (Naciones Unidas, 2020a). En paralelo a esta transformación, tuvo lugar la diversificación de orígenes que llegan a los diferentes contextos de destino latinoamericanos, incidiendo sobre la tradicional dinámica migratoria fronteriza dominante en buena parte de ellos (Prieto Rosas y Zapata, 2023; Vera y Adler, 2020). Esta transformación

se refleja en el crecimiento de la población de origen haitiano en Chile, peruano y colombiano en Argentina, o cubano y dominicano en Uruguay (Bengochea y Pellegrino, 2023; Prieto Rosas y Montiel, 2020; Cerruti y Parrado, 2015).

A las transformaciones anteriormente mencionadas se le adiciona, en 2020, el inicio de la pandemia por COVID-19, la cual demarcó consecuencias específicas sobre la movilidad internacional a nivel mundial y específicamente en los flujos intrarregionales latinoamericanos (Cerruti, 2022). En particular, el cierre de fronteras decretado en los países latinoamericanos durante los primeros meses de 2020 dio lugar a un período caracterizado por la inmovilidad involuntaria, que afectó a aquellas personas que planeaban emigrar, pero también a quienes se encontraban en tránsito y a personas con necesidad de protección internacional (Garcés Mascareñas y López-Sala, 2023; Cerruti, 2022). Las restricciones a movilidad y la existencia de múltiples situaciones de vulnerabilidad en los países de destino, dieron lugar a un fenómeno novedoso para el contexto latinoamericano, que refiere a procesos masivos de retorno hacia el país de origen, los cuales tuvieron como contrapartida la existencia de importantes contingentes de personas migrantes varadas en frontera (Mcauliffe et al., 2021; (In)movilidad en las Américas, s.f.). A este período signado por la paralización de la movilidad le siguió una paulatina reactivación de los desplazamientos con características particulares, como el incremento de la participación de personas migrantes bajo situaciones de mayor vulnerabilidad y solicitantes de protección internacional, así como una tendencia creciente de la movilidad hacia terceros países con un fuerte componente de irregularidad documental especialmente en aquellos flujos con dirección norte (Prieto Rosas et al., en prensa; OIM, 2022b; CEPAL, 2020)

Por otro lado, es importante mencionar cuatro aspectos que se han vuelto una parte constitutiva de la movilidad intrarregional latinoamericana en el período reciente. Primero, se destaca la creciente naturaleza mixta de los flujos, conformada por la migración de tipo laboral y las personas solicitantes de protección internacional (Prieto Rosas y Zapata, 2023). El incremento de personas con necesidad de protección internacional responde al aumento de desplazados venezolanos y a la llegada de personas migrantes de origen caribeño a países de América del Sur, ya sea con el propósito de establecerse en dicho destino o tomándolo como un eslabón en el tránsito hacia Estados Unidos (Prieto Rosas y Zapata, 2023; Gandini, Fernández de la Reguera y Narváez Gutiérrez, 2020). Probablemente, uno de los casos más sobresalientes en términos del incremento relativo de solicitudes de la condición de refugiado alude a la población cubana en Uruguay, la cual alcanzó las 10853 solicitudes en 2019, lo que representó un aumento del 3000% respecto a 2016 (Prieto Rosas y Montiel, 2020).

Segundo, se ha identificado un paulatino crecimiento de la participación femenina en los flujos intrarregionales en América Latina, aunque dicha tendencia difiere según países de destino, encontrando que, por ejemplo, Argentina y Uruguay presentan un perfil feminizado para la población migrante, mientras lo opuesto sucede en Brasil (Stefoni, 2018). También se presentan heterogeneidades en la composición por sexo entre orígenes de un mismo país de destino (Giorguli et al., 2023; Prieto Rosas et al., 2022).

Tercero, como fue mencionado anteriormente, los itinerarios migratorios intrarregionales han ganado en complejidad en tanto la movilidad se da crecientemente de manera escalonada y con viajes migratorios más prolongados, de manera tal que la prevalencia de la movilidad a un tercer país se incrementó (Prieto Rosas et al., en prensa; OIM, 2022b). Incluso, la reorientación de los flujos y movilidad desde terceros países se vio acrecentada con posterioridad al momento crítico de la pandemia por COVID-19 que estuvo signado por el cierre de fronteras y la paralización de trámites administrativos que posibilitan la regularización documental (Bengochea et al., 2022; OIM, 2022b).

Cuarto, como un elemento novedoso pero propio de la movilidad internacional en el período reciente, se destaca la relevancia de las redes sociales web como espacio de divulgación de información y organización por parte de las poblaciones migrantes a lo largo del viaje migratorio (Morán, 2022; Gandini, Fernández de la Reguera y Narváez Gutiérrez, 2020; Melella y Perret, 2016). Ejemplo claro de ello, es el uso que le han dado las “caravanas” de migrantes centroamericanos en tránsito por territorio mexicano para el desarrollo de estrategias de movilidad (Gandini, Fernández de la Reguera y Narváez Gutiérrez, 2020). Asimismo, las redes sociales web han jugado un rol relevante en clave de generar espacios de vinculación en el país de destino con otros connacionales o bien, con familiares o amigos que permanecen en el país de origen (Melella y Perret, 2016). Consecuentemente, un aspecto de relevancia a considerar es el acceso a internet por parte de las poblaciones migrantes en el contexto de destino, frente a lo cual Varona (2022) identifica que para el caso mexicano es variable según sexo y país de nacimiento. En particular, las mujeres y las personas nacidas en países sudamericanos tienen un mayor acceso a internet en detrimento de los varones y los nacionales de países centroamericanos, así como también se observa una incidencia más elevada entre los migrantes absolutos respecto a los recientes (Varona, 2022).

Objetivos de investigación

Objetivo general

Analizar en qué medida los datos de usuarios *expats* de Facebook se aproximan a la cuantificación de poblaciones migrantes que arrojan las encuestas de hogares en países latinoamericanos en 2021 y 2022.

Objetivos específicos

- Sistematizar la información disponible sobre migrantes internacionales en las encuestas de hogares de Latinoamérica y el Caribe entre 2020 y 2022.
- Estimar los stocks de usuarios de Facebook etiquetados como “anteriormente viviendo en...” para cada díada origen-destino y sexo a partir de descargas diarias de información de Facebook Ads Manager.
- Estimar los sesgos en los datos de usuarios de Facebook anteriormente viviendo en otro país con respecto a las cifras de stocks de migrantes disponibles en las encuestas de hogares en 2021 y 2022 según la definición de migrante adoptada (absolutos o recientes).
- Comparar la magnitud de las diferencias observadas entre los datos de usuarios de Facebook y personas migrantes captadas por las encuestas de hogares con aquellas observables en la comparación de datos de usuarios de Facebook con las cifras de stocks de migrantes de Naciones Unidas.
- Analizar si existen diferencias por sexo en los sesgos de los datos de usuarios de Facebook anteriormente viviendo en otro país con respecto a las cifras de stocks de migrantes disponibles en las encuestas de hogares y en estimaciones de Naciones Unidas.

Datos y metodología

Descripción de los datos

Con el propósito de dar respuesta a los objetivos de investigación planteados se trabajó de manera conjunta con tres fuentes de información: (i) cifras de usuarios diarios de Facebook (DAU)⁶ etiquetados como “anteriormente viviendo en...” que proporciona la aplicación de Facebook Ads Manager a través de su Interfaz de Programación de Aplicaciones (API, por sus siglas en inglés) versión 16.0 a 18.0, (ii) microdatos de encuestas de hogares, y (iii) estimaciones del *International Migrant Stock* 2020 de Naciones Unidas.

Datos de Facebook

Desde diciembre de 2019, en el marco del proyecto de investigación “Using internet-based data to quantify and sample international migrants. Applications to examine recent immigration to Uruguay” (Programa de Población, Universidad de la República y el Max Planck Institute for Demographic Research, Alemania), se realizan consultas diarias sobre las cifras de MAU y DAU de Facebook para 273 díadas origen-destino latinoamericanas (21 países de destino y 13 países de origen), desagregadas por sexo y grupos de edad (18-24, 25-54, 55-65 años) e identificadas a partir del lugar de residencia actual y anterior (Facebook nombra a esta variable como “Han vivido en j (anteriormente expatriados - j)”) disponible en la segmentación por comportamientos que provee Facebook Ads Manager.

Entre los países de destino para los que Facebook cuenta con información, y que fueron utilizados en este trabajo, están Argentina, Bolivia Brasil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, Guatemala, Haití, Honduras, México, Nicaragua, Panamá, Paraguay, Perú, Puerto Rico, República Dominicana, Trinidad y Tobago, Uruguay y Venezuela. En tanto, los orígenes nacionales latinoamericanos y caribeños etiquetados como “anteriormente viviendo en...” para los que Facebook provee cifras de audiencias son Argentina, Brasil, Chile, Colombia, Cuba, El Salvador, Guatemala, Haití, Honduras, México, Nicaragua, Perú, Puerto Rico, República Dominicana y Venezuela. En las consultas diarias que se realizan desde 2019 se extrajo cifras de MAU y DAU para todos los orígenes mencionados excepto México y Puerto Rico. Es importante mencionar que Facebook no publica información sobre orígenes de

⁶ Siguiendo la puntualización realizada por Gil-Clavel, Zagheni y Bordone (2022) y Prieto Rosas et al. (2022b), se optó por trabajar con el DAU en lugar del MAU, ya que el segundo invisibiliza la magnitud de grupos poblacionales de tamaño reducido como es el caso de algunas comunidades de origen dentro de la población migrante.

relevancia para la caracterización de la dinámica migratoria intrarregional latinoamericana como Bolivia y Paraguay, lo cual implica que, por ejemplo, para el caso de Argentina, no se incluya en el análisis a dos de los principales países de origen que componen el stock de personas nacidas en el extranjero.

En particular, para este trabajo se utilizó el conjunto de descargas diarias de información, es decir se estimó una cifra sintética anual del DAU según año, origen, destino y sexo. Para ello, se programó un procedimiento *Bootstrap* que consistió en seleccionar de manera aleatoria una observación para cada mes según año, origen, destino y sexo y se calculó la mediana anual con base en las 12 observaciones seleccionadas (Gil-Clavel, Zagheni y Bordone, 2022). Dicho proceso fue repetido 2500 veces, siguiendo la sugerencia de Efron y Tibshirani (1994) y para obtener el valor final del stock se calculó la media de las 2500 medianas de cifras de DAU para cada combinación de año, origen, destino y sexo. De este modo, se obtuvo una medida sintética que busca contemplar la varianza en los valores del DAU a nivel anual, así como también permitió estimar los correspondientes intervalos de confianza.

Cabe señalar que no se encuentra disponible documentación asociada a los elementos considerados por la plataforma para la construcción de la etiqueta “Han vivido en *j* (anteriormente expatriados - *j*)”. No obstante, a partir de lo reportado por investigadores/as que trabajan en Facebook, se ha sostenido que existen dos elementos de relevancia que sirven de insumo: la información autoreportada por los usuarios sobre su ciudad natal y las características de las redes de amigos en Facebook (Rampazzo et al., 2021; Spyrtos et al., 2019; Herdağdelen et al., 2016). Por otro lado, estimaciones brindadas por Facebook muestran que para el año 2018 había alrededor de 11% de cuentas duplicadas y 5% de cuentas falsas de usuarios mensuales a nivel mundial, siendo mayor la incidencia entre mercados en desarrollo (U.S. SEC, 2019). No obstante, en esta oportunidad no fueron corregidas las cifras de usuarios de Facebook según este factor por falta de un corrector desagregado por origen y destino.

Además de brindar la posibilidad de monitorear la magnitud de las cifras de *expats* en tiempo real, Facebook es relevante en tanto representa la principal red social en buena parte de los países de América Latina (Gráfico 7 en Anexo) de acuerdo con datos del Latinobarómetro (2020). Como ya fue señalado en la sección precedente, las mujeres suelen tener un mayor nivel de uso de la red social respecto a los varones, aunque dicha tendencia no es homogénea entre países (Gráfico 7 en Anexo). Específicamente, para más de la mitad de los contextos considerados, las mujeres cuentan con un porcentaje de uso mayor (Argentina, Brasil, Chile, Costa Rica, Paraguay, Perú, República Dominicana y Uruguay), mientras que para un grupo no menor la distribución tiende a ser simétrica (Colombia, Ecuador, Honduras, México,

Nicaragua y Venezuela) y se identifica una mayor prevalencia entre los varones para el caso de El Salvador (Gráfico 7 en Anexo).

Por otro lado, un factor que puede estar incidiendo en la prevalencia del uso de Facebook y que fue considerado en algunos de los análisis de los sesgos entre datos tradicionales y digitales (Varona, 2022; Palotti et al., 2020; Spyrtos et al., 2019), es el nivel de penetración de internet en los hogares. En el caso de América Latina, se puede observar que el acceso es heterogéneo, siendo los países del Cono Sur y Costa Rica los que tienen una mayor participación de personas con conexión a internet en el hogar, en detrimento de ciertos países centroamericanos, como Nicaragua y Guatemala que no superan el umbral del 25% (Gráfico 8 en Anexo). Como señalan Palotti et al. (2020), estas desiguales estructuras de acceso a internet inciden sobre el uso que hacen las poblaciones migrantes de las redes sociales como Facebook y, por ende, repercuten en los sesgos de selección que afectan las magnitudes y composición de las audiencias reportadas.

Encuestas de hogares

A partir de la revisión de formularios y microdatos de las encuestas de hogares, es posible clasificar a los países de destino seleccionados según tres aspectos: i) existencia de una encuesta multipropósito de referencia para el período de interés (2020⁷-2022); ii) accesibilidad a microdatos, formularios y diccionarios de variables; y iii) disponibilidad de preguntas que permitan identificar a las personas migrantes y su origen nacional (país). Por consiguiente, de los 21 países de destino para los que se cuenta con información de Facebook, se identificaron nueve que cumplen con los tres criterios mencionados: Argentina, Chile, Colombia, Costa Rica, Ecuador, México, Paraguay, Perú y Uruguay (Cuadro 6 en Anexo). Entre los países que no cumplen con el primer criterio se encuentran Haití y Venezuela; en función del segundo criterio se descartaron los casos de Nicaragua, Panamá, Puerto Rico y Trinidad y Tobago; y por no satisfacer una o dos de las condiciones definidas en el tercero de los criterios se excluyeron Bolivia, Brasil, El Salvador, Guatemala, Honduras y República Dominicana (Cuadro 6 en Anexo). En síntesis, las encuestas de hogares consideradas en el análisis son: Encuesta Permanente de Hogares (EPH) (Argentina), Encuesta Nacional de Empleo (ENE) (Chile), Gran Encuesta Integrada de Hogares (GEIH) (Colombia), Encuesta Nacional de Hogares (ENH) (Costa Rica), Encuesta Nacional de Empleo, Subempleo y Desempleo

⁷ Cabe señalar que, si bien se tienen datos de audiencias de Facebook desde 2019, el año 2020 es el primero para el que se dispone de información anual completa.

(ENEMDU) (Ecuador), Encuesta Nacional de Ocupación y Empleo (ENOE) (México), Encuesta Permanente de Hogares (EPH) (Paraguay), Encuesta Nacional de Hogares (ENAHO) (Perú) y Encuesta Continua de Hogares (ECH) (Uruguay). Por otra parte, si bien se sistematizó la información para el trienio 2020-2022, el estudio se acota a los años 2021 y 2022 debido a que en los formularios de las encuestas de Ecuador y Uruguay no se incluyeron las preguntas que permiten captar la movilidad internacional en 2020 (Cuadro 6 en Anexo).

Si bien las encuestas de referencia para los nueve países de destino que componen el universo de estudio comparten la ventaja de poder identificar a las personas migrantes, la pregunta que permite cumplir con dicho objetivo no es homogénea. Específicamente, Argentina, Colombia, Ecuador, México y Uruguay indagan sobre el lugar de nacimiento de la persona encuestada, mientras que Chile, Costa Rica, Paraguay y Perú consultan por el lugar de residencia de la madre al momento del nacimiento del/de la encuestado/a. De este modo, a partir de la combinación del lugar de nacimiento declarado por la persona encuestada y el de residencia actual -que equivale al país donde se realizó el levantamiento de la información-, es posible identificar a los migrantes absolutos, es decir, aquellos que residen en un lugar diferente al de nacimiento sin distinción de la fecha del último viaje migratorio. Adicionalmente, si se conoce el lugar de residencia de la persona encuestada en un momento anterior al del levantamiento de la información, es posible conocer qué subconjunto de los migrantes absolutos son migrantes recientes, categoría que alude generalmente a aquellas personas que llevan hasta cinco años de residencia en el país.

El Cuadro 1 contiene la información por encuesta sobre la disponibilidad de preguntas para identificar migrantes absolutos y recientes para cada origen nacional considerado en el análisis. A modo de ejemplo, vemos que en el caso de Colombia en 2021 (Cuadro 1.a) es posible identificar migrantes absolutos y recientes nacidos en Argentina, Perú y Venezuela, mientras que en 2022 se encuentra disponible la información para los doce orígenes considerados (Cuadro 1.b). En términos generales, todos los países, excepto Ecuador, disponen de una pregunta que permite identificar a migrantes recientes (Cuadro 1), aunque hay algunas diferencias en el horizonte temporal considerado: i) Argentina, Chile, Colombia, Perú, Paraguay y Uruguay indagan sobre el lugar de residencia cinco años antes; ii) Costa Rica lo hace en referencia a los dos años previos; y iii) México pregunta sobre el lugar de residencia un año antes del levantamiento de la información. Además, otra dimensión que impone diferencias entre encuestas nacionales refiere a la desagregación por país de origen, oscilando entre la especificación exhaustiva para los dos años de interés (Argentina, Chile, Ecuador,

Una puntualización adicional tiene que ver con que la migración reciente también se podría haber analizado según el país de procedencia (Varona, 2022), pero dado que en Uruguay hay problemas de calidad en el relevamiento de la variable para 2022, y que, en el resto de los países, en más del 90% de los casos hay una coincidencia entre país de nacimiento y país de residencia previo, no es viable considerar este análisis dado que está trabajando con muestras.

Cabe señalar que la periodicidad para la que se encuentran publicados los microdatos de las encuestas de hogares estableció una diferencia en el proceso de estimación de los stocks migratorios por díadas origen-destino y sexo entre países de destino. Específicamente, la información para Chile, Costa Rica, Ecuador, Paraguay, Perú y Uruguay se presenta en bases anualizadas, mientras que en el caso de Argentina y México se encuentran desagregadas por trimestre y para Colombia por meses. Por consiguiente, para los tres últimos países se calculó el promedio de los stocks de los trimestres o meses según año, origen y sexo con el propósito de tener un valor resumen anual.

Estimaciones de Naciones Unidas

En el caso de las estimaciones de stock de migrantes internacionales que publica Naciones Unidas, se tomó como referencia la actualización de 2020 ya que es el año más reciente para el que se han publicado los datos. Las cifras aluden a migrantes absolutos y en el caso de los países de destino seleccionados -excepto Chile y Colombia- son resultado de la combinación de la proyección demográfica con base en la población nacida en el exterior proveniente de datos censales y encuestas y registros administrativos de personas que se encuentran bajo la condición de refugiado, son solicitantes de asilo o desplazados venezolanos según ACNUR (Naciones Unidas, 2020b). En el caso de Chile y Colombia, solo se considera el primero de los componentes (Naciones Unidas, 2020b). Asimismo, se establecen diferencias con relación a la desagregación por origen nacional que se publica para cada país de destino (Cuadro 2).

Cuadro 2. Disponibilidad de cifras de stock de migrantes absolutos según díadas origen-destino en *International Migration Stock* (Naciones Unidas), 2020

| | | País de destino | | | | | | | | | |
|-------------------------------------|-----|-----------------|-----|-----|-----|-----|-----|-----|-----|-----|--|
| | | ARG | CHL | COL | CRI | ECU | MEX | PRY | PER | URY | |
| País de origen (país de nacimiento) | ARG | ■ | X | X | X | X | X | X | X | X | |
| | BRA | X | X | X | X | X | X | X | X | X | |
| | CHL | X | ■ | X | X | X | X | X | X | X | |
| | COL | X | X | ■ | X | X | X | ■ | X | ■ | |
| | CUB | X | X | X | X | X | X | ■ | X | ■ | |
| | SLV | X | ■ | X | X | X | X | ■ | X | ■ | |
| | GTM | X | ■ | X | X | X | X | ■ | X | ■ | |
| | HTI | X | X | ■ | X | X | X | ■ | X | ■ | |
| | HND | X | ■ | X | X | X | X | ■ | X | ■ | |
| | NIC | X | ■ | ■ | X | X | X | ■ | X | ■ | |
| | PER | X | X | X | X | X | X | ■ | ■ | X | |
| | DOM | X | X | X | X | X | X | ■ | X | ■ | |
| | VEN | X | X | X | X | X | X | ■ | X | X | |

X Estimación disponible para el origen nacional
 ■ Estimación no disponible para el origen nacional
 ■ No aplica

Nota: los códigos de países refieren a la clasificación ISO3.

Fuente: elaboración propia a partir de *International Migrant Stock 2020* – Naciones Unidas.

Cabe señalar que las estimaciones según origen nacional se presentan desagregadas por sexo, mientras que la distribución por grupos quinquenales de edad se encuentra disponible únicamente para el total de los migrantes en el país de destino. Dado que en los extractos diarios de Facebook programados se tiene únicamente información para los usuarios de 18 a 65 años, se imputó un factor de corrección a las estimaciones de Naciones Unidas por díadas origen-destino y sexo con el propósito de obtener una magnitud del stock de migrantes comparable entre fuentes. Para el cálculo de dicho factor de corrección se calculó la proporción de población de 20 a 64 años⁹ según sexo (Tabla 1), la cual fue imputada a las observaciones según origen-destino-sexo. De este modo, se intentó aproximar una corrección de la magnitud del grupo poblacional de estudio para cada origen, destino y sexo a partir de la estructura etaria del conjunto general de migrantes por destino y sexo.

⁹ Dado que Naciones Unidas no publica los stocks de migrantes desagregados por edades simples, el grupo de 20 a 64 años debe ser tratado como un proxy de la población de 18 a 65 años.

Tabla 1. Factor de corrección (proporción de población de 20 a 64 años) de los stocks de migrantes provistos por Naciones Unidas según destino y sexo, 2020

| País de destino | Varones | Mujeres |
|-----------------|---------|---------|
| Argentina | 0.32 | 0.31 |
| Chile | 0.16 | 0.18 |
| Colombia | 0.42 | 0.43 |
| Costa Rica | 0.25 | 0.22 |
| Ecuador | 0.45 | 0.46 |
| México | 0.65 | 0.65 |
| Paraguay | 0.27 | 0.28 |
| Perú | 0.38 | 0.40 |
| Uruguay | 0.47 | 0.47 |

Fuente: elaboración propia a partir de *International Migrant Stock 2020*.

Es de relevancia recordar que la utilización de los datos de Naciones Unidas en el marco del presente trabajo tiene un fin netamente comparativo y es utilizado como insumo para valorar el grado de ajuste entre Facebook y las encuestas de hogares, así como en qué medida difieren los sesgos de Facebook respecto a las encuestas y las estimaciones de Naciones Unidas.

La comparación de las magnitudes de los stocks estimadas muestra que en general las cifras derivadas de encuestas de hogares se aproximan a las de usuarios diarios de Facebook, donde estas últimas suelen ser más elevadas que su estándar de comparación (Gráfico 1.a y 1.b). Sin embargo, se encuentran algunas excepciones notables como el caso de los nacionales de Nicaragua residentes en Costa Rica o las personas nacidas en Perú captadas en Chile (Gráfico 1.a y 1.b). Otro aspecto a destacar refiere a la amplitud de los intervalos de confianza estimados para las encuestas de hogares, lo que muestra síntomas de debilidad de dicha fuente para proporcionar estimaciones precisas sobre los stocks según origen nacional. De hecho, en casos como el de Argentina y México se observa con claridad la superposición entre los intervalos estimados para las encuestas y Facebook, lo que significa que no existen diferencias significativas en la magnitud de los stocks entre fuentes a un 95% de confianza (Gráfico 1.a y 1.b). Por otro lado, la distribución de las observaciones estimadas para Facebook y Naciones Unidas en 2020 muestra una tendencia general similar a la descrita anteriormente, aunque con diferencias entre destinos identificadas con claridad para el caso mexicano y chile (Gráfico 1.c.).

Gráfico 1. Número absoluto de migrantes según fuente de información, definición de migrante empleada y país de origen por año. Países de destino y origen seleccionados, 2020-2022

Gráfico 1.a. Información para el año 2021

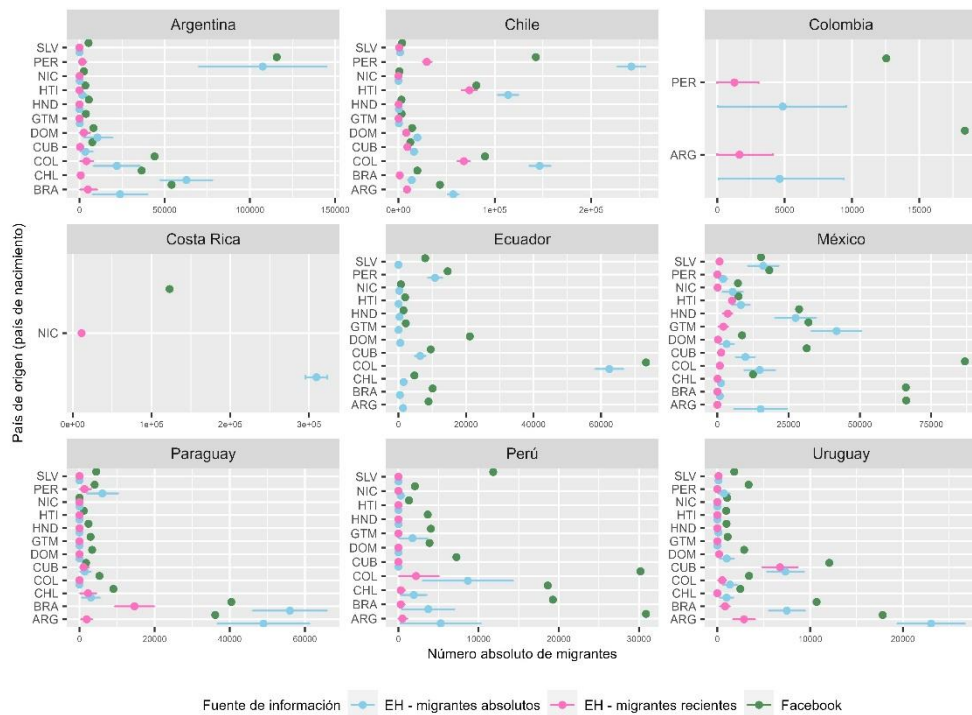


Gráfico 1.b. Información para el año 2022

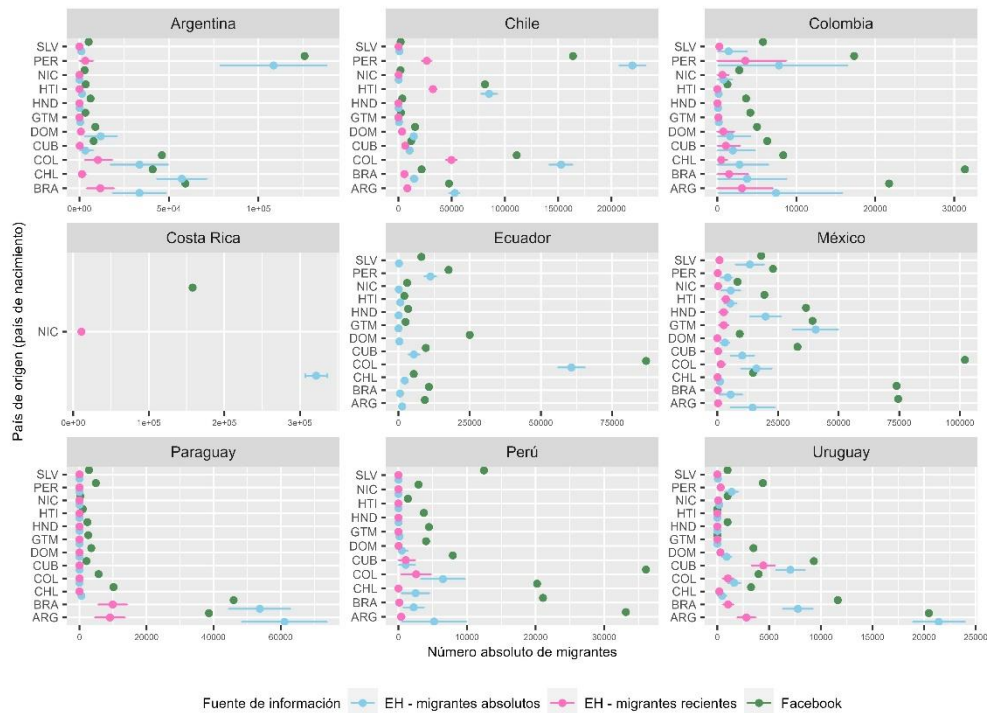
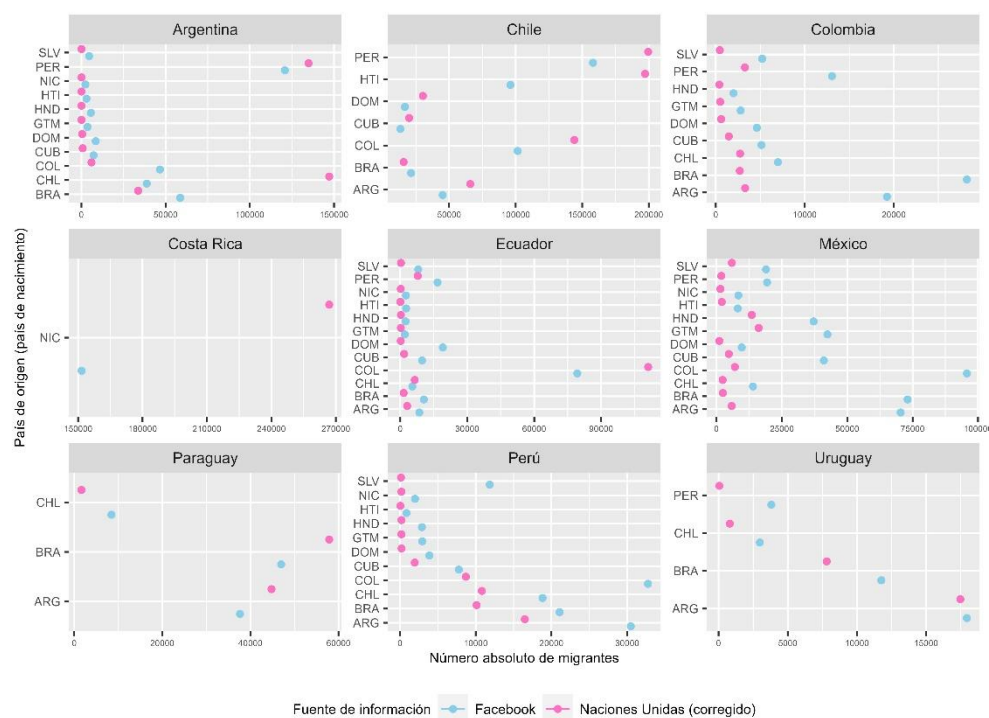


Gráfico 1.c. Información para el año 2020



Nota: intervalos de confianza a un 95% de confianza (excepto para las estimaciones de Naciones Unidas que no tienen un intervalo de confianza asociado). Las escalas de los ejes no son comparables ya que la magnitud del stock para cada día asume valores muy diferentes. Para el cálculo de los intervalos de confianza de las encuestas de Argentina, México y Colombia se utilizó el promedio de los intervalos por trimestre en el caso de los dos primeros países, y por mes en el caso de Colombia. Asimismo, en los casos que el límite inferior del intervalo estimado con base en la encuesta de hogar asumía un valor negativo, se imputó el valor cero dado que no es posible consignar un stock poblacional menor a cero. Con el propósito de favorecer la visualización se omitió la presentación de las estimaciones para Venezuela.

Fuente: elaboración propia a partir de cifras de usuarios activos diarios de Facebook, microdatos de la EPH (Argentina), ENE (Chile), GEIH (Colombia), ENH (Costa Rica), ENEMDU (Ecuador), ENOE (México), EPH (Paraguay), ENAHO (Perú) y ECH (Uruguay) y estimaciones de *International Migrant Stock 2020* (Naciones Unidas).

A modo de síntesis, y retomando diversos aspectos mencionados hasta el momento sobre las fuentes de información empleadas, se presenta un cuadro resumen sobre las principales virtudes y limitaciones de cada una, destacando aquellas que afectan de manera directa los objetivos de investigación planteados:

Cuadro 3. Síntesis de las limitaciones y fortalezas* de los datos de Facebook, encuestas de hogares y estimaciones de Naciones Unidas para la medición de la migración internacional en América Latina

| | Fortalezas | Limitaciones |
|-----------------------------|---|---|
| Facebook | <ul style="list-style-type: none"> • Extracción y almacenamiento de datos en tiempo real • Datos recolectados con periodicidad diaria (posibilidad de controlar shocks externos y cambios en los comportamientos demográficos) • Aproximación a poblaciones de “difícil acceso” u ocultas | <ul style="list-style-type: none"> • Representatividad • Existencia de sesgos de selección • Nulo control del investigador sobre cambios en los algoritmos de construcción de las variables provistas por Facebook Ads Manager • Falta de documentación sobre la construcción de variables y el proceso de generación de la información (i.e. son Black-Box) |
| Encuestas de hogares | <ul style="list-style-type: none"> • Periodicidad de recolección anual • Identificación de migrantes según país de nacimiento y país de residencia en un horizonte temporal previo al levantamiento de la información • Disponibilidad de documentación y metadatos sobre el proceso de generación de la información • Inclusión de migrantes en situación documental irregular | <ul style="list-style-type: none"> • Restricciones derivadas del tamaño muestral cuando se trabaja con comunidades de origen pequeñas • Subestimación de las magnitudes de los stocks de migrantes por a) exclusión de viviendas colectivas; b) uso de ponderadores construidos con proyecciones de población que pueden estar desactualizadas; y/o c) eventual desactualización del marco muestral |
| Naciones Unidas | <ul style="list-style-type: none"> • Disponibilidad de información sobre migración internacional para la totalidad de los países de destino • Disponibilidad de documentación y metadatos sobre el proceso de generación de la información | <ul style="list-style-type: none"> • Diferencias en la medición de la migración internacional entre países de destino • Incompleta desagregación por origen nacional para ciertos países de destino • Falta de desagregación por edad de las estimaciones de stocks de migrantes por díadas origen-destino |

Nota: (*) se realizó una selección de las limitaciones y fortalezas priorizando aquellas que se vinculan directamente con los objetivos de la tesis.

Fuente: elaboración propia a partir de metadatos de encuestas de hogares y Naciones Unidas; Perdomo Rico (2022); Rampazzo et al. (2021); Gutiérrez et al. (2020); Prieto Rosas y Márquez Scotti, (2019), Alburez et al. (2019).

Unidad de análisis

A partir de la descripción de los datos empleados para el análisis y los objetivos específicos de investigación definidos, las unidades de análisis, entendidas como las díadas origen-destino, se agrupan en dos conjuntos, los cuales se describen a continuación.

El primer grupo responde al objetivo que busca estimar los sesgos según la definición de migrante adoptada –absolutos vs. recientes– y está conformado por ocho de los países de destino seleccionados. Se excluye a Ecuador por no contar con información sobre migrantes recientes. En síntesis, se tiene un total de 79 díadas para 2021 y 88 para 2022 (Cuadro 4).

Cuadro 4. Díadas origen-destino que conforman la unidad de análisis para la estimación de los sesgos según definición de migrante. Países de destino y origen seleccionados

| | | País de destino | | | | | | | |
|-------------------------------------|------|-----------------|-----|-----|-----|-----|-----|-----|-----|
| | | ARG | CHL | COL | CRC | MEX | PRY | PER | URY |
| País de origen (país de nacimiento) | ARG | | X | X | | X | X | X | X |
| | BRA | X | X | X* | | X | X | X | X |
| | CHL | X | | X* | | X | X | X | X |
| | COL | X | X | | | X | X | X | X |
| | CUB | X | X | X* | | X | X | X | X |
| | SLV | X | X | X* | | X | X | X | X |
| | GTM | X | X | X* | | X | X | X | X |
| | HTI | X | X | X* | | X | X | X | X |
| | HND | X | X | X* | | X | X | X | X |
| | NIC | X | X | X* | X | X | X | X | X |
| | PER | X | X | X | | X | X | | X |
| | DOM | X | X | X* | | X | X | X | X |
| | VEN | X | X | X | | X | X | X | X |
| Total de orígenes | 2021 | 12 | 12 | 3 | 1 | 13 | 13 | 12 | 13 |
| | 2022 | 12 | 12 | 12 | 1 | 13 | 13 | 12 | 13 |

X Estimación del stock para la díada en 2021 y 2022

X* Estimación del stock para la díada únicamente en 2022

Estimación no disponible para el origen nacional

No aplica

Nota: los códigos de países refieren a la clasificación ISO3.

Fuente: elaboración propia.

El segundo grupo se ajusta al propósito de comparar los sesgos de Facebook respecto a las encuestas de hogares y Naciones Unidas para migrantes absolutos, por lo que se contempla a los nueve países de destino seleccionados. De este modo, el conjunto de díadas comparable difiere según el año de referencia de las encuestas de hogares que se seleccione, siendo mayor el número de observaciones para 2022 –70 en 2021 y 77 en 2022– (Cuadro 5).

Cuadro 5. Díadas origen-destino que conforman la unidad de análisis para la estimación de los sesgos de Facebook respecto a las encuestas de hogares y Naciones Unidas. Países de destino y origen seleccionados

| | | País de destino | | | | | | | | |
|-------------------------------------|-------------------------|-----------------|-----|-----|-----|-----|-----|-----|-----|-----|
| | | ARG | CHL | COL | CRC | ECU | MEX | PRY | PER | URY |
| País de origen (país de nacimiento) | ARG | ■ | X | X | ■ | X | X | X | X | X |
| | BRA | X | X | X* | ■ | X | X | X | X | X |
| | CHL | X | ■ | X* | ■ | X | X | X | X | X |
| | COL | X | X | ■ | ■ | X | X | ■ | X | ■ |
| | CUB | X | X | X* | ■ | X | X | ■ | X | ■ |
| | SLV | X | ■ | X* | ■ | X | X | ■ | X | ■ |
| | GTM | X | ■ | X* | ■ | X | X | ■ | X | ■ |
| | HTI | X | X | ■ | ■ | X | X | ■ | X | ■ |
| | HND | X | ■ | X* | ■ | X | X | ■ | X | ■ |
| | NIC | X | ■ | ■ | X | X | X | ■ | X | ■ |
| | PER | X | X | X | ■ | X | X | ■ | ■ | X |
| | DOM | X | X | X* | ■ | X | X | ■ | X | ■ |
| | VEN | X | X | X | ■ | X | X | ■ | X | X |
| Total de orígenes | 2020 (UN) vs. 2021 (EH) | 12 | 8 | 3 | 1 | 13 | 13 | 3 | 12 | 5 |
| | 2020 (UN) vs. 2022 (EH) | 12 | 8 | 10 | 1 | 13 | 13 | 3 | 12 | 5 |

- X Estimación del stock comparable para la día en 2021 y 2022
- X* Estimación del stock comparable para la día únicamente en 2022
- Estimación comparable no disponible para el origen nacional
- No aplica

Nota: los códigos de países refieren a la clasificación ISO3.

Fuente: elaboración propia.

Metodología

Replicando la metodología empleada por Zagheni, Weber y Gummedi (2017), se buscó dar respuesta a los objetivos de investigación de medición de las estimaciones de stock de migrantes según distintas fuentes por medio de la estimación de una serie de modelos lineales multivariados que permitieron cuantificar los sesgos. Los modelos se organizan en función de los dos grupos de unidades de análisis descritos anteriormente: estimación de los sesgos según definición de migrante (Ecuación 1 y 2) y comparación de la dirección y magnitud de los sesgos de Facebook respecto a las encuestas de hogares y Naciones Unidas para migrantes absolutos (Ecuación 3 a 5). De este modo, la estrategia a implementar se sustenta bajo la óptica de métodos de calibración con presencia de datos “reales” (Zagheni y Weber, 2015), donde se busca conocer la correlación entre éstos y los datos de usuarios de Facebook.

Es preciso señalar una serie de decisiones metodológicas que se tomaron en el proceso de especificación de los modelos. En primer lugar, en lo que respecta a las encuestas de hogares y los modelos estimados para los dos conjuntos de unidades de análisis, se optó por especificar modelos separados para cada año de referencia. Esta decisión responde a dos principales motivos que se vinculan a la pandemia por COVID-19: por un lado, las restricciones en materia sanitaria que afectaron los procesos de recolección de información (CEPAL, 2022) y, por otro lado, el cierre y posterior progresiva apertura de fronteras nacionales que afectó la movilidad internacional, y fundamentalmente a aquellas personas que se encontraban en tránsito (Bengochea et al., 2022). En segundo lugar, siguiendo la lógica de *dummy* y *contrast coding* (UCLA, s.f.) se generaron una serie de variables *dummy* para cada país de destino y origen, donde la categoría de referencia refiere al resto de los destinos y orígenes considerados en el análisis según corresponda. Consecuentemente, en la estimación de los modelos se tuvo que omitir a un país de destino y uno de origen para evitar el problema de colinealidad perfecta. De este modo, la selección de México como el país de destino¹⁰ omitido alude a que posee información para los 13 orígenes seleccionados en todos los modelos, mientras que en lo que respecta al origen se decidió que el país omitido fuera Venezuela dado que es el país de nacimiento que conforma una diáda con el mayor número de destinos para la totalidad de las ecuaciones. En tercer lugar, la razón que sustenta el uso del logaritmo en la variable dependiente y uno de los regresores, es la necesidad de contar con magnitudes comparables de los stocks entre las diádas origen-destino no afectadas por el tamaño de la población general de cada diáspora. Asimismo, al trabajar con los logaritmos y dado que las cifras de stock de Facebook y las encuestas de hogares pueden asumir el valor cero, se adicionó una unidad a cada una de las observaciones, con el propósito de evitar el sesgo de exclusión de comunidades de origen pequeñas resultante de la imposibilidad de calcular el logaritmo de cero.

La ecuación 1 y 2 aluden a los modelos de estimación de los sesgos de Facebook respecto a las encuestas de hogares según la definición de migrante considerada: absolutos y recientes:

Ecuación 1:

$$\log(\text{stockabseh}_{ij}^k) = \beta_0 + \beta_1 \log(\text{stockfbk}_{ij}^k) + \beta_2(\text{destino1}) + \dots + \beta_9(\text{destino8}) + \beta_{10}(\text{origen1}) + \dots + \beta_{23}(\text{origen13}) + \varepsilon_{ij}^k,$$

donde *i* refiere al origen, *j* al destino y *k* al año de observación.

¹⁰ Inicialmente se intentó estimar los sesgos de manera individual para cada uno de los países de destino, pero, producto del número de observaciones o la existencia de importantes niveles de multicolinealidad aproximada, se optó por especificar un modelo general que tenga como control al país de destino.

Ecuación 2:

$$\log(\text{stockreche}_{ij}^k) = \beta_0 + \beta_1 \log(\text{stockfbk}_{ij}^k) + \beta_2(\text{destino1}) + \dots + \beta_9(\text{destino8}) + \beta_{10}(\text{origen1}) + \dots + \beta_{23}(\text{origen13}) + \varepsilon_{ij}^k,$$

donde i refiere al origen, j al destino y k al año de observación.

Cuadro 6. Especificación de las variables utilizadas en la ecuación 1 y 2

| Variable | Nombre | Especificación |
|--------------------------|--|---|
| stockabseh _{ij} | Número absoluto de migrantes de origen i que residen en el país destino j según la encuesta de hogar | Continua, con valores positivos mayores a cero |
| stockreche _{ij} | Número absoluto de migrantes recientes de origen i que residen en el país destino j según la encuesta de hogar | Continua, con valores positivos mayores a cero |
| stockfbk _{ij} | Mediana anual de los usuarios diarios de Facebook en j que residían anteriormente en i | Continua, con valores positivos mayores a cero |
| origen _i | País de origen | Categorías: origen _{n-i} (referencia); origen _i |
| destino _i | País de destino | Categorías: destino _{n-i} (referencia); destino _i |

Fuente: elaboración propia.

Las ecuaciones 3 a 5 refieren a los modelos para la estimación y comparación de los sesgos de Facebook respecto a las encuestas de hogares y Naciones Unidas:

Ecuación 3:

$$\log(\text{stockun}_{ijk}^{2020}) = \beta_0 + \beta_1 \log(\text{stockfbk}_{ijk}^{2020}) + \beta_2(\text{destino1}) + \dots + \beta_{10}(\text{destino9}) + \beta_{11}(\text{origen1}) + \dots + \beta_{24}(\text{origen13}) + \beta_{25}\text{sexo} + \varepsilon_{ijk}^{2020},$$

donde i refiere al origen, j al destino y k al sexo.

Ecuación 4:

$$\log(\text{stockabseh}_{ijk}^{2021}) = \beta_0 + \beta_1 \log(\text{stockfbk}_{ijk}^{2021}) + \beta_2(\text{destino1}) + \dots + \beta_{10}(\text{destino9}) + \beta_{11}(\text{origen1}) + \dots + \beta_{24}(\text{origen13}) + \beta_{25}\text{sexo} + \varepsilon_{ijk}^{2021},$$

donde i refiere al origen, j al destino y k al sexo.

Ecuación 5:

$$\log(\text{stockabseh}_{ijk}^{2022}) = \beta_0 + \beta_1 \log(\text{stockfbk}_{ijk}^{2022}) + \beta_2(\text{destino1}) + \dots + \beta_{10}(\text{destino9}) + \beta_{11}(\text{origen1}) + \dots + \beta_{24}(\text{origen13}) + \beta_{25}\text{sexo} + \varepsilon_{ijk}^{2022},$$

donde i refiere al origen, j al destino y k al sexo.

Cuadro 7. Especificación de las variables utilizadas en las ecuaciones 3 a 5

| Variable | Nombre | Especificación |
|---------------------------|--|---|
| stockabseh _{ijk} | Número absoluto de migrantes de origen <i>i</i> que residen en el país destino <i>j</i> del sexo <i>k</i> según la encuesta de hogar | Continua, con valores positivos mayores a cero |
| stockun _{ijk} | Número absoluto de migrantes de origen <i>i</i> que residen en el país destino <i>j</i> del sexo <i>k</i> según Naciones Unidas | Continua, con valores positivos mayores a cero |
| stockfb _{kj} | Mediana anual de los usuarios diarios de Facebook en <i>j</i> que residían anteriormente en <i>i</i> del sexo <i>k</i> | Continua, con valores positivos mayores a cero |
| Sexo | Sexo | Categorías: varón (referencia); mujer |
| origen _i | País de origen | Categorías: origen _{n-i} (referencia); origen _i |
| destino _i | País de destino | Categorías: destino _{n-i} (referencia); destino _i |

Fuente: elaboración propia.

Análisis de los sesgos de los datos de Facebook respecto a las encuestas de hogares para migrantes absolutos y recientes

El presente apartado de resultados tiene como propósito valorar a cuál definición de migrante –absolutos o recientes– aplicada a los datos de encuestas de hogares se asemeja más el número de usuarios de Facebook etiquetados como “anteriormente viviendo en...”. Para ello, se trabaja con siete países de destino de América Latina para los que es posible establecer la comparación con los datos disponibles de Facebook: Argentina, Chile, Colombia, Costa Rica, México, Paraguay, Perú y Uruguay. A continuación, se presenta la evaluación gráfica de la relación entre ambas fuentes de información y los resultados de los modelos descritos en la ecuación 1 y 2, lo cual permite tener una aproximación al nivel de ajuste entre ambas fuentes de información. Esta estrategia analítica, también permite apreciar la magnitud de la sobre o subrepresentación de las audiencias diarias de Facebook respecto a las encuestas para cada año de estimación controlando por el país de destino y origen.

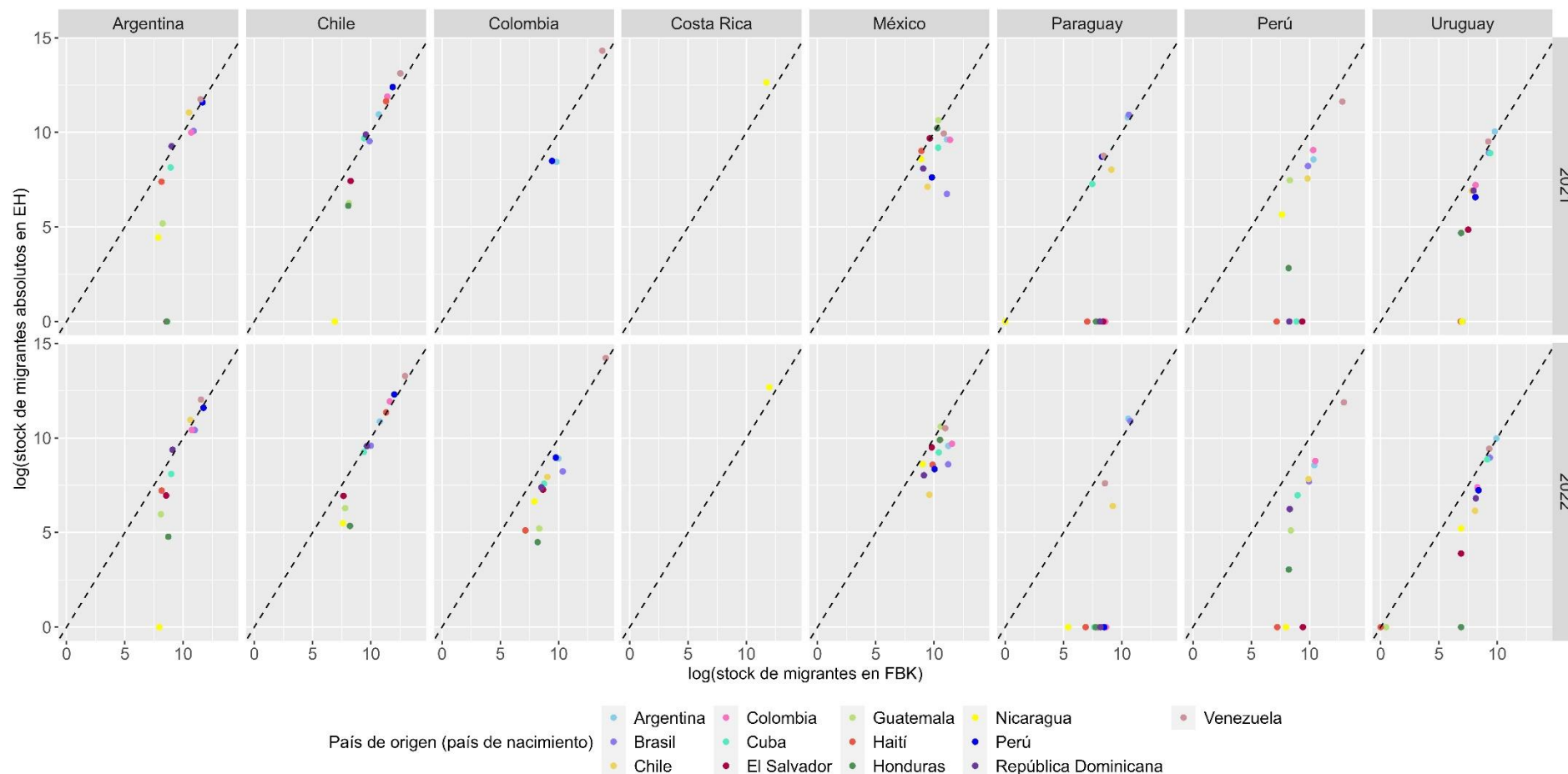
En primer lugar, la comparación gráfica para migrantes absolutos, muestra que en general, los puntos se aproximan a la recta de 45 grados, lo que implica que el valor de la magnitud de los stocks de Facebook y las encuestas de hogares son semejantes (Gráfico 3). No obstante, se identifican diferencias de ajuste por país de destino y origen considerados. La dispersión de los puntos en torno a la recta de referencia parece mostrar a grandes rasgos que, para ambos años de observación, la semejanza entre fuentes es mayor para destinos como

Argentina, Chile y Costa Rica, y en segundo lugar para Colombia y México (Gráfico 3). Asimismo, en la mayoría de los destinos se aprecia una alineación de la mayoría de los puntos sobre la función de identidad con sobresalientes excepciones para ciertos orígenes, lo cual se evidencia con claridad en el caso de Uruguay, Perú y Paraguay, y en menor medida en Argentina, Chile y Colombia (Gráfico 2). Estos *outliers* comparten dos aspectos comunes: por un lado, una tendencia que muestra patrones de sobrerrepresentación de Facebook respecto a las encuestas de hogares, y por otro lado, el tratarse de orígenes muy poco numerosos en los destinos mencionados, pertenecientes a países de Centroamérica y el Caribe. Otro aspecto a destacar refiere a que para la amplia mayoría de los orígenes de relevancia en la dinámica migratoria de los destinos seleccionados se evidencia una fuerte semejanza entre los valores de stock que reportan ambas fuentes cuando se considera tanto a migrantes recientes como a aquellos con más tiempo de asentamiento en el país de residencia actual. Con relación a los años de estimación, no se aprecia un patrón claro de diferencias en la visualización gráfica.

En segundo lugar, la evaluación gráfica para migrantes recientes muestra un comportamiento general a la sobrerrepresentación de los stocks en Facebook, incluso para aquellos países de destino que mostraban un buen nivel de ajuste para la definición abordada anteriormente (Gráfico 3). Cabe señalar que esta característica se presenta incluso para contextos que tienen diferencias en el horizonte temporal considerado para la identificación de migrantes recientes. En este sentido, una primera aproximación gráfica parece mostrar que la comparación respecto a las encuestas de hogares muestra una mayor semejanza para los migrantes absolutos, lo cual va en la misma línea que lo encontrado por Varona (2022). Respecto a la comparación entre años de relevamiento, se observa la misma tendencia que para los migrantes absolutos, es decir, no se observan notables diferencias según año (Gráfico 3). No obstante, una importante modificación respecto a la definición precedente refiere al aumento de observaciones atípicas asociadas a países que cuentan con un stock casi nulo en las encuestas de hogares, en conjunto con la diversificación de los orígenes que las componen (Gráfico 3). Este crecimiento es previsible si se considera que los migrantes recientes son un subgrupo de los migrantes absolutos, y por ende, necesariamente tendrán una magnitud del stock menor. Por otro lado, vale la pena destacar la similitud en el número estimado de migrantes venezolanos entre ambas fuentes de datos, siendo uno de los pocos casos que presenta un muy buen grado de ajuste para las dos definiciones y años de referencia mencionados, en conjunto con los segundos y terceros orígenes de relevancia para cada contexto de destino (Gráfico 2 y 3). De este modo, una primera aproximación gráfica parece mostrar que la definición que más se aproxima a la cuantificación de audiencias de Facebook

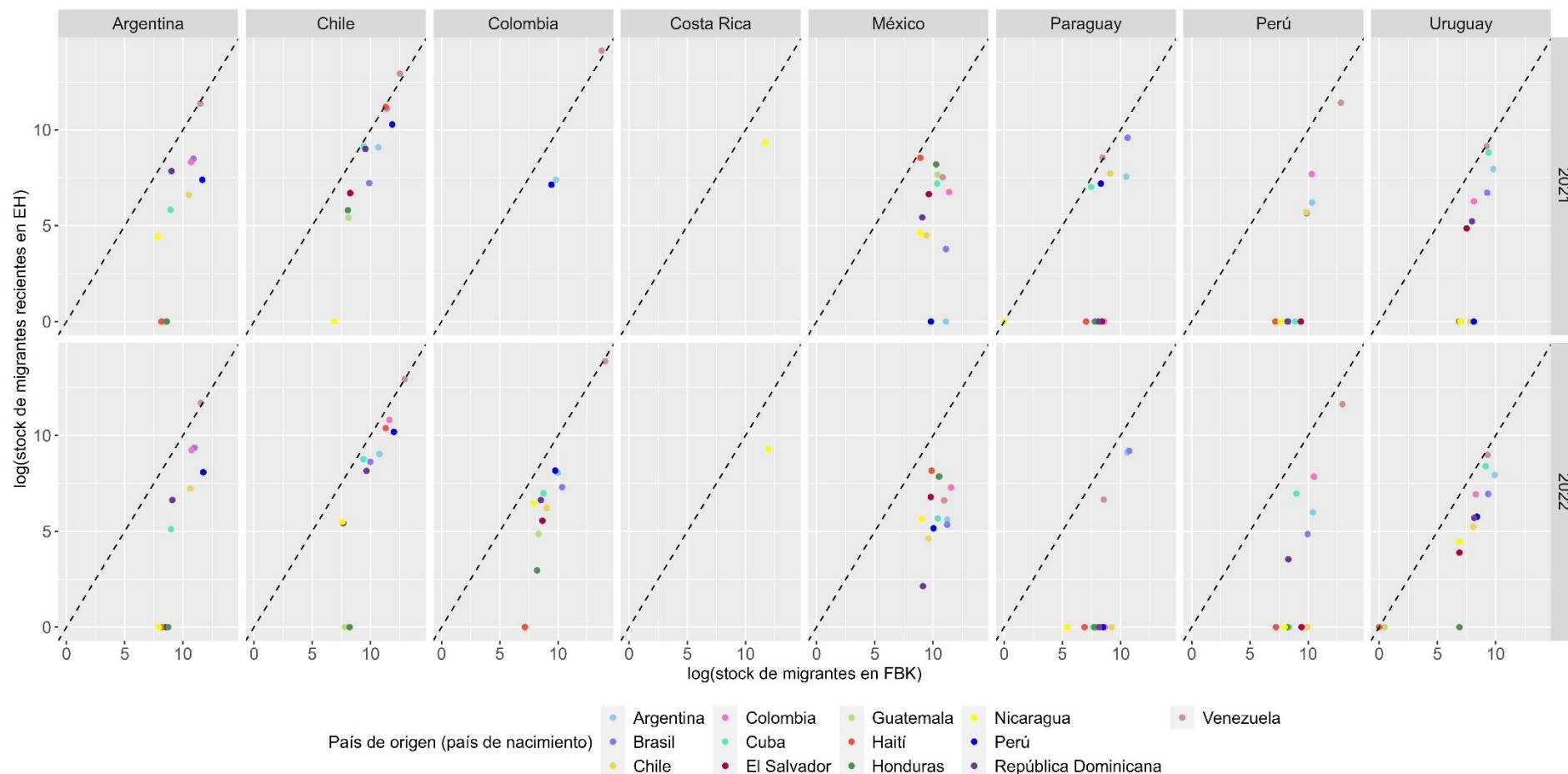
es la de migrantes absolutos, pero con variabilidad en el grado de semejanza entre las fuentes según el país de destino y origen considerado.

Gráfico 2. Relación entre el stock de migrantes del conjunto de datos de Facebook y las estimaciones de migrantes absolutos de encuestas de hogares según país de origen y destino por año. Países seleccionados, 2021 y 2022



Fuente: elaboración propia a partir de cifras de usuarios activos diarios de Facebook y microdatos de la EPH (Argentina), ENE (Chile), GEIH (Colombia), ENH (Costa Rica), ENOE (México), EPH (Paraguay), ENAHO (Perú) y ECH (Uruguay).

Gráfico 3. Relación entre el stock de migrantes del conjunto de datos de Facebook y las estimaciones de migrantes recientes de encuestas de hogares según país de origen y destino por año. Países seleccionados, 2021 y 2022



Fuente: elaboración propia a partir de cifras de usuarios activos diarios de Facebook y microdatos de la EPH (Argentina), ENE (Chile), GEIH (Colombia), ENH (Costa Rica), ENOE (México), EPH (Paraguay), ENAHO (Perú) y ECH (Uruguay).

Las Tablas 2 y 3 muestran los resultados de la estimación de los modelos lineales multivariados descritos en la ecuación 1 y 2 según año de relevamiento. En particular, la información presentada brinda insumos sobre el nivel de ajuste entre fuentes de información a partir del valor del estadístico R^2 ajustado, así como una estimación de los sesgos entre Facebook y las encuestas de hogares por medio de los coeficientes de cada país de destino y origen. El signo positivo de los coeficientes indica que Facebook está subrepresentando el stock de migrantes respecto a las encuestas de hogares y el signo negativo lo inverso.

De este modo, la evaluación analítica de la correlación entre fuentes confirma lo hallado por medio de la aproximación gráfica, en clave de la adecuación de cada una de las definiciones. Específicamente, el valor del R^2 ajustado de los modelos lineales multivariados estimados es sistemáticamente mayor en el caso de aquellos que refieren a la definición de migrante absoluto respecto a los que emplean la definición de migrante reciente (Tabla 2 y 3). Además, la adecuación entre fuentes de información se incrementa notoriamente para el segundo año de estudio, estableciéndose un aumento de veinte puntos porcentuales para migrantes absolutos y dieciséis para migrantes recientes respecto a 2021 (Tabla 2 y 3). Otro aspecto relevante a destacar es que la estimación progresiva de los modelos permitió observar que los coeficientes de los países de destino varían una vez que se incluye como control a los países de origen, lo cual implica que ambos grupos de regresores no son independientes (Tabla 2 y 3).

En términos de los sesgos según origen, para aquellos coeficientes que son estadísticamente significativos a un 95% de confianza, se identifica una tendencia general a la sobreestimación de los datos de Facebook (Tabla 2 y 3), lo cual marca una clara diferencia con los resultados alcanzados por Zagheni, Weber y Gummadi (2017), quienes encontraron un comportamiento inverso respecto a la *American Community Survey*. En cambio, dichos resultados coinciden con lo identificado por Rosati, Cerruti y Maguire (2021) para el caso de nacionales venezolanos en la Ciudad Autónoma de Buenos Aires y Spyrtos et al., (2019) para las personas de origen latinoamericano residentes en Estados Unidos. Por ejemplo, el coeficiente de El Salvador para migrantes absolutos en 2021 muestra que Facebook sobrerrepresenta sistemáticamente el stock de migrantes de dicho origen respecto al resto de orígenes, manteniendo constante el país de destino y el logaritmo de los usuarios *expats* de Facebook. Con relación a los sesgos según país de destino, se observa que la magnitud de la brecha entre fuentes parece ser menos relevante que la identificada para los países de origen cuando se controla por estos últimos (Tabla 2 y 3). Asimismo, las diferencias según país de destino parecen ser más relevantes para la definición de migrante absoluto en 2022 y

dependiendo del año y definición se identifican algunas disparidades destacables (Tabla 3). Por ejemplo, para Perú, en ambos años de estimación y para la definición de migrante absoluto, se encuentra que Facebook sobrerrepresenta sistemáticamente al conjunto de orígenes considerados en ese destino cuando se toma como referencia al grupo de los restantes destinos y se controla por el origen (Tabla 2 y 3). En cambio, para Chile en 2021 y Uruguay en 2022 se identifica que Facebook subrepresenta al conjunto de migrantes recientes manteniendo constante el origen (Tabla 2 y 3).

Tabla 2. Coeficientes del modelo de regresión lineal del logaritmo del stock de **migrantes absolutos y recientes** de origen i en el destino j con base en las encuestas de hogares. Países seleccionados, **2021**

| | log(stock migrantes absolutos EH) | | | log(stock migrantes recientes EH) | | |
|-------------------------|-----------------------------------|---------------------|---------------------|-----------------------------------|---------------------|---------------------|
| | (1) | (2) | (3) | (1) | (2) | (3) |
| log(<i>expats</i> FBK) | 1.638*** (0.304) | 1.534*** (0.365) | 1.004* (0.398) | 1.482*** (0.326) | 1.492*** (0.374) | 1.096* (0.432) |
| País de destino | | | | | | |
| Argentina | | -0.702 (0.964) | -0.84 (1.095) | | 0.374 (1.222) | 0.142 (1.221) |
| Chile | | 0.539 (0.62) | 0.523 (1.052) | | 3.092** (1.011) | 3.018* (1.246) |
| Colombia | | 0.0851 (0.596) | -1.247 (0.984) | | 2.745** (1.015) | 2.45 (1.522) |
| Costa Rica | | 1.208 (0.749) | 3.013 (2.108) | | 1.443 (1.119) | 2.423 (1.822) |
| Paraguay | | -1.342 (1.272) | -2.516 (1.404) | | 1.517 (1.392) | 0.64 (1.572) |
| Perú | | -2.567* (1.042) | -2.893* (1.254) | | -1.149 (1.168) | -1.564 (1.288) |
| Uruguay | | -0.113 (1.039) | -1.178 (1.356) | | 1.318 (1.392) | 0.522 (1.541) |
| País de origen | | | | | | |
| Argentina | | | -0.565 (0.938) | | | -3.440* (1.586) |
| Brasil | | | -1.202 (1.283) | | | -2.407 (1.212) |
| Chile | | | -0.869 (1.147) | | | -2.879 (1.5) |
| Colombia | | | -2.122 (1.374) | | | -2.418 (1.447) |
| Cuba | | | -1.874 (1.574) | | | -1.673 (1.492) |
| El Salvador | | | -4.960** (1.838) | | | -4.482* (1.793) |
| Guatemala | | | -3.409 (1.918) | | | -5.022** (1.747) |
| Honduras | | | -4.325* (1.842) | | | -4.828* (1.839) |
| Haití | | | -3.560* (1.775) | | | -3.802 (1.93) |
| Nicaragua | | | -3.249 (2.238) | | | -3.538 (1.864) |
| Perú | | | -0.913 (1.059) | | | -4.208** (1.5) |
| R. Dominicana | | | -2.968 (1.752) | | | -2.974 (1.757) |
| Constante | -8.136** (2.962) | -6.54 (3.697) | 1.114 (4.939) | -8.568** (3.119) | -9.602* (3.854) | -2.399 (5.347) |
| N | 79 | 79 | 79 | 79 | 79 | 79 |
| R ² | 0.528 | 0.586 | 0.669 | 0.452 | 0.56 | 0.65 |
| R ² ajustado | 0.522 | 0.539 | 0.555 | 0.445 | 0.509 | 0.529 |

Nota: * p<0.05, ** p<0.01, *** p<0.001. Error estándar robusto a heteroscedasticidad entre paréntesis. Se empleó *dummy* y *contrast coding*, y para evitar el problema de especificación de colinealidad perfecta se omitió a México como destino y Venezuela como origen.

Fuente: elaboración propia a partir de cifras de usuarios activos diarios de Facebook y microdatos de la EPH (Argentina), ENE (Chile), GEIH (Colombia), ENH (Costa Rica), ENOE (México), EPH (Paraguay), ENAHO (Perú) y ECH (Uruguay).

Tabla 3. Coeficientes del modelo de regresión lineal del logaritmo del stock de **migrantes absolutos y recientes** de origen *i* en el destino *j* con base en las encuestas de hogares. Países seleccionados, **2022**

| | log(stock migrantes absolutos EH) | | | log(stock migrantes recientes EH) | | |
|-------------------------|-----------------------------------|---------------------|----------------------|-----------------------------------|----------------------|----------------------|
| | (1) | (2) | (3) | (1) | (2) | (3) |
| log(<i>expats</i> FBK) | 1.439*** (0.214) | 1.407*** (0.173) | 1.048*** (0.173) | 1.291*** (0.237) | 1.467*** (0.221) | 1.060*** (0.252) |
| País de destino | | | | | | |
| Argentina | | 0.0441 (0.66) | -0.0736 (0.768) | | -0.251 (0.924) | -0.448 (1.043) |
| Chile | | 0.805* (0.337) | 0.728 (0.594) | | 1.997* (0.785) | 1.74 (0.949) |
| Colombia | | 0.077 (0.409) | -0.303 (0.657) | | 1.933** (0.645) | 1.586 (0.922) |
| Costa Rica | | 1.234** (0.401) | 3.401* (1.333) | | 0.786 (0.571) | 1.665 (1.323) |
| Paraguay | | -3.522** (1.059) | -4.239*** (1.079) | | -1.21 (1.014) | -2.022 (1.126) |
| Perú | | -2.259** (0.804) | -2.608** (0.869) | | -1.274 (0.925) | -1.667 (0.977) |
| Uruguay | | 1.142 (0.611) | -0.0191 (0.799) | | 3.627*** (0.744) | 2.313* (0.935) |
| País de origen | | | | | | |
| Argentina | | | -0.28 (1.155) | | | -1.783 (1.211) |
| Brasil | | | -0.928 (1.131) | | | -1.803 (1.225) |
| Chile | | | -1.136 (0.956) | | | -4.023** (1.265) |
| Colombia | | | -1.831 (0.981) | | | -1.772 (1.219) |
| Cuba | | | -1.593 (0.89) | | | -1.787 (1.206) |
| El Salvador | | | -3.147* (1.272) | | | -4.017** (1.359) |
| Guatemala | | | -2.236 (1.139) | | | -4.174** (1.536) |
| Honduras | | | -4.085*** (1.178) | | | -5.504*** (1.538) |
| Haití | | | -2.231* (1.074) | | | -3.203 (1.694) |
| Nicaragua | | | -3.456* (1.461) | | | -3.030* (1.472) |
| Perú | | | -2.022* (0.9) | | | -2.951** (-1) |
| R. Dominicana | | | -1.625 (0.996) | | | -2.744* (1.275) |
| Constante | -6.209** (2.101) | -5.394** (1.797) | 0.194 (2.406) | -6.750** (2.288) | -9.044*** (2.325) | -2.029 (3.363) |
| N | 88 | 88 | 88 | 88 | 88 | 88 |
| R ² | 0.583 | 0.752 | 0.812 | 0.476 | 0.667 | 0.76 |
| R ² ajustado | 0.578 | 0.727 | 0.756 | 0.47 | 0.634 | 0.689 |

Nota: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Error estándar robusto a heteroscedasticidad entre paréntesis. Se empleó *dummy* y *contrast coding*, y para evitar el problema de especificación de colinealidad perfecta se omitió a México como destino y Venezuela como origen.

Fuente: elaboración propia a partir de cifras de usuarios activos diarios de Facebook y microdatos de la EPH (Argentina), ENE (Chile), GEIH (Colombia), ENH (Costa Rica), ENOE (México), EPH (Paraguay), ENAHO (Perú) y ECH (Uruguay).

En síntesis, es posible jerarquizar los resultados en dos líneas: i) una tendencia general a la sobrerrepresentación del número de migrantes por díadas origen-destino en Facebook respecto a las encuestas de hogares y ii) un mayor nivel de correlación entre ambas fuentes de información para migrantes absolutos. Respecto al primer punto, es posible esbozar la hipótesis de que la sobreestimación de los datos digitales respecto al estándar de comparación responde a la conjugación de características propias de cada una de las fuentes: el diseño y tamaño muestral de las encuestas, así como la eventual desactualización de los marcos muestrales (Perdomo Rico, 2022; Gutiérrez et al., 2020), y la incidencia de cuentas duplicadas o falsas en Facebook, que a su vez tiende a ser mayor en países de residencia en desarrollo (Rampazzo et al., 2021; U.S. SEC, 2019; Gil-Clavel y Zagheni, 2019). Como se retomará en la sección “Conclusiones”, este último aspecto puede ser uno de los factores que esté explicando los resultados opuestos encontrados respecto a Zagheni, Weber y Gummadi (2017). Por otro lado, el hecho de encontrar una correlación más elevada entre fuentes para migrantes absolutos es esperable si se considera que los migrantes recientes son un subconjunto de aquellos que residen en el país de destino sin distinción de la fecha de llegada. De este modo, si se identifica una tendencia general a la sobrerrepresentación de los datos de Facebook para migrantes absolutos, dicha brecha será mayor para los recientes.

En otro orden, la aproximación gráfica y analítica permiten vislumbrar dos resultados emergentes. Primero, se encuentra que el ajuste entre fuentes es muy elevado para los orígenes nacionales más numerosos de cada país de destino para las dos definiciones de migrante y años considerados, lo cual permite considerar a Facebook como una fuente de relevancia para predecir en tiempo real los stocks de migrantes en la región latinoamericana tomando como referencia los factores de corrección derivados de la estimación de los sesgos respecto a las encuestas de hogares. Segundo, las encuestas de hogares parecen mostrar importantes restricciones para captar la magnitud de orígenes nacionales con muy baja representación, lo cual lleva a plantear la idea de que los datos de audiencias de Facebook pueden comportarse como una fuente de información complementaria para aproximar la magnitud de dichas comunidades nacionales.

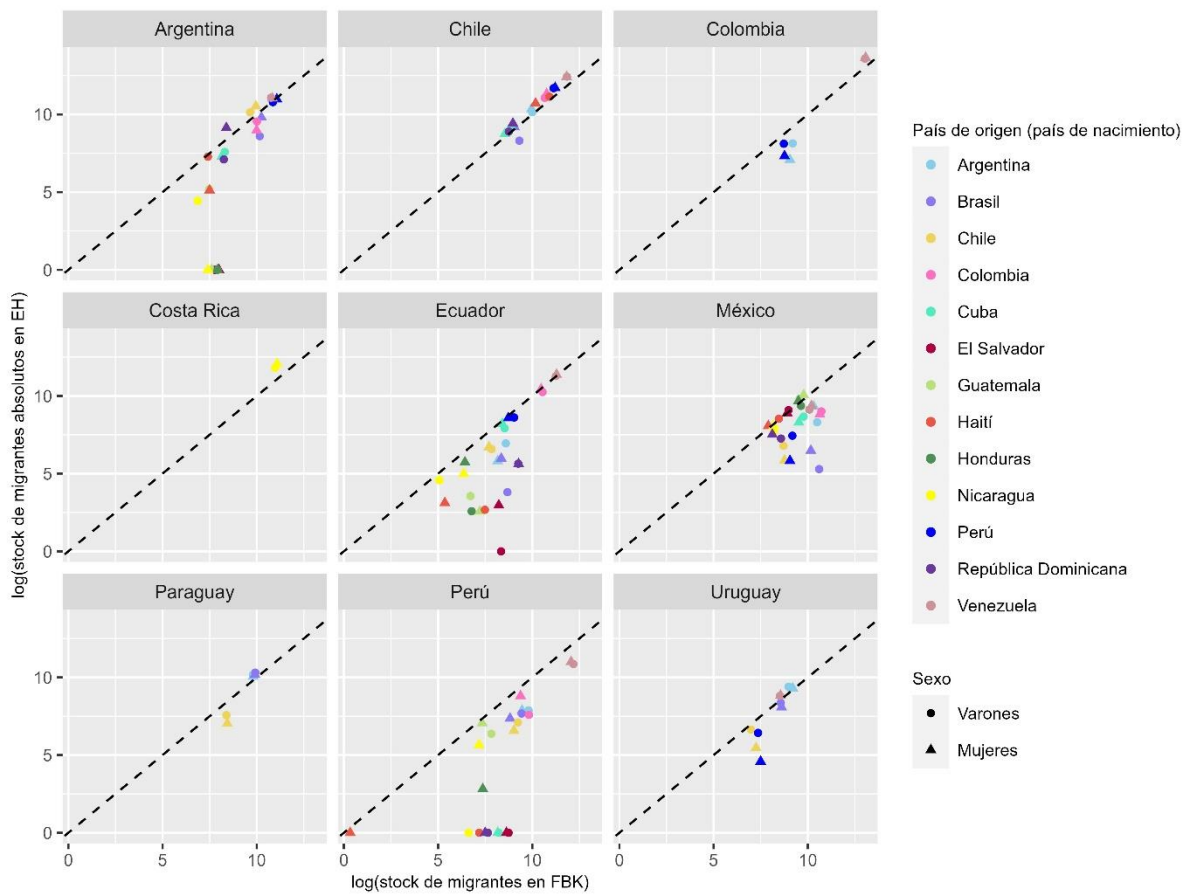
Comparación de los sesgos de los datos de Facebook respecto a las encuestas de hogares y estimaciones de Naciones Unidas

La presente sección de resultados tiene como objetivo analizar la correlación entre los datos de Facebook y las encuestas de hogares en función de la comparación con el nivel de ajuste de los primeros respecto a las estimaciones de Naciones Unidas, y observar si existen diferencias

según sexo. Al igual que en el apartado precedente, se abordará el propósito por medio del análisis gráfico y de los modelos multivariados descritos en las ecuaciones 3 a 5 en el capítulo de “Datos y Metodología”. Cabe recordar que, a diferencia del análisis presentado para las definiciones de migrante, en esta oportunidad el análisis se limita a la definición de migrante absoluto ya que es la que es comparable con las estimaciones de Naciones Unidas, y que las unidades de análisis están conformadas por nueve países de destino: Argentina, Chile, Colombia, Costa Rica, Ecuador, México, Paraguay, Perú y Uruguay.

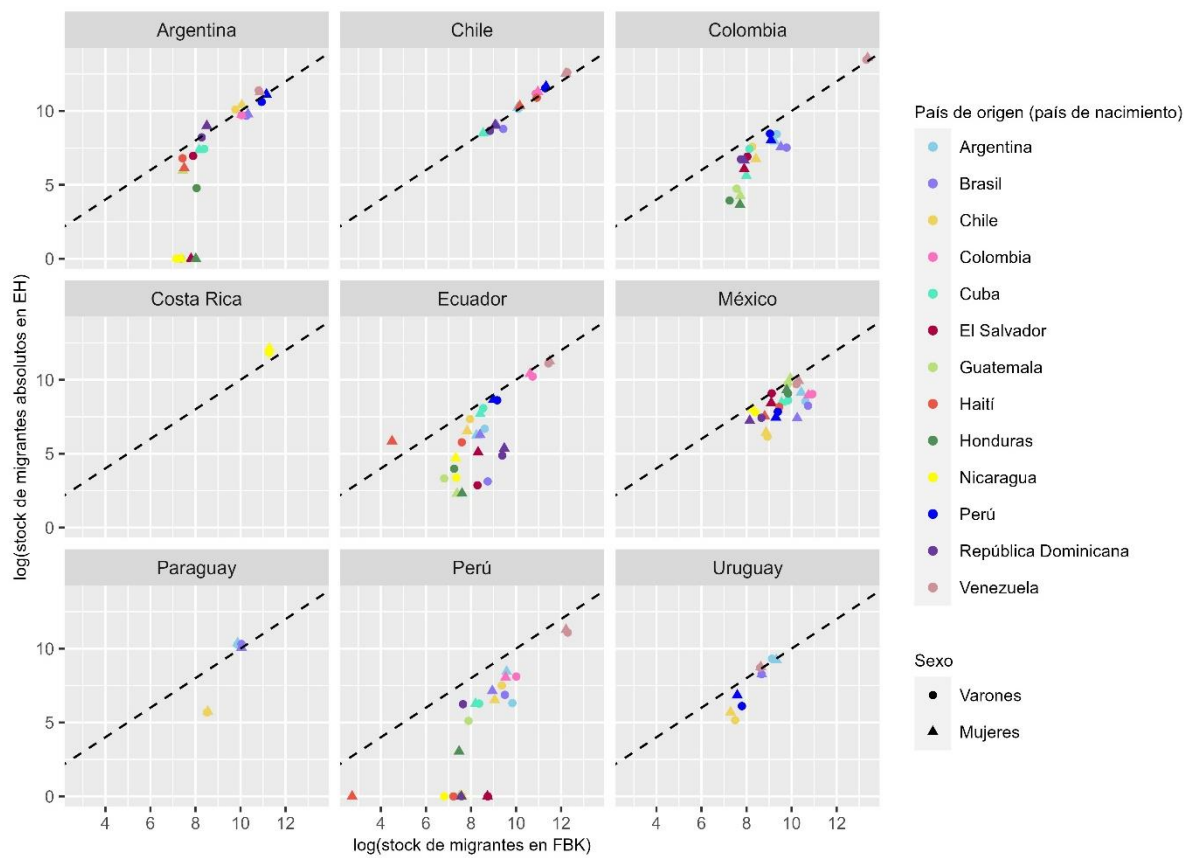
Uno de los principales resultados alcanzados refiere a la presencia de semejanzas en las tendencias de (des)ajuste de los datos de Facebook respecto a las encuestas de hogares y a las estimaciones de Naciones Unidas. Por ejemplo, en países como Colombia, Ecuador, México y Perú los datos digitales muestran en general un valor superior a sus dos estándares de comparación, y una distribución semejante de la posición de cada origen (Gráfico 4 a 6). Sin embargo, existen algunas excepciones a lo mencionado, como el caso de Argentina o Uruguay, para el cual las estimaciones de Facebook respecto a las de Naciones Unidas presentan una mayor dispersión en torno a la recta de referencia si se lo compara con la correlación entre Facebook y las encuestas de hogares (Gráfico 4 a 6). Asimismo, una diferencia adicional que se establece entre las fuentes tomadas como referencia para la evaluación de los datos de Facebook refiere a que en la comparación realizada con datos de Naciones Unidas hay una menor presencia de *outliers* para ciertos países como Argentina y Perú (Gráfico 5 y 6), aspecto que reafirma la idea planteada anteriormente sobre las limitaciones que presentan las encuestas de hogares para proporcionar una estimación precisa del stock de comunidades de origen muy pequeñas. Respecto a las distribuciones según sexo, en términos generales en la revisión gráfica no se observan diferencias sustantivas entre fuentes (Gráfico 4 a 6), lo que lleva a plantear la hipótesis de que el sexo no segmenta de manera clara la correlación entre Facebook y cada una de las fuentes de referencia en la comparación.

Gráfico 4. Relación entre el stock de migrantes del conjunto de datos de Facebook y las estimaciones de migrantes absolutos de encuestas de hogares según país de origen y destino por sexo. Países seleccionados, 2021



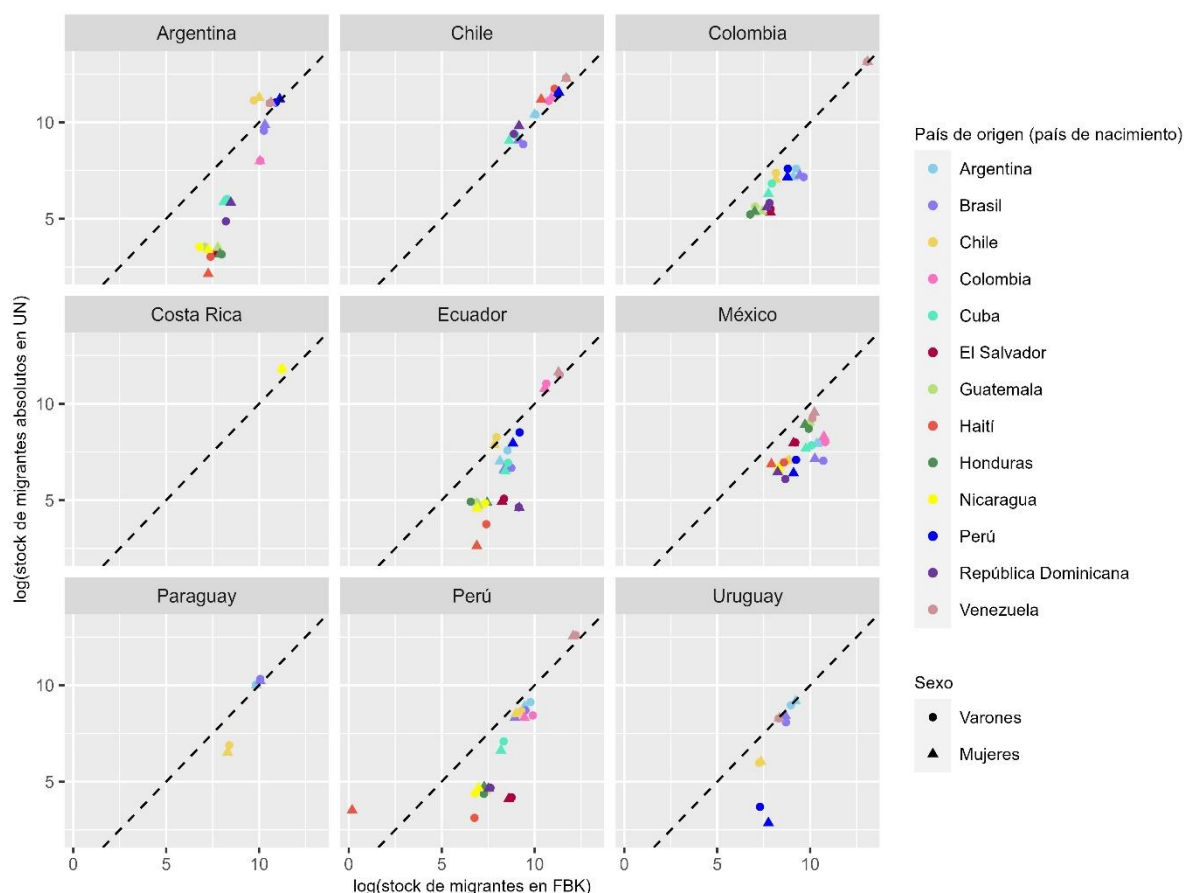
Fuente: elaboración propia a partir de cifras de usuarios activos diarios de Facebook y microdatos de la EPH (Argentina), ENE (Chile), GEIH (Colombia), ENH (Costa Rica), ENEMDU (Ecuador), ENOE (México), EPH (Paraguay), ENAHO (Perú) y ECH (Uruguay).

Gráfico 5. Relación entre el stock de migrantes del conjunto de datos de Facebook y las estimaciones de migrantes absolutos de encuestas de hogares según país de origen y destino por sexo. Países seleccionados, 2022



Fuente: elaboración propia a partir de cifras de usuarios activos diarios de Facebook y microdatos de la EPH (Argentina), ENE (Chile), GEIH (Colombia), ENH (Costa Rica), ENEMDU (Ecuador), ENOE (México), EPH (Paraguay), ENAHO (Perú) y ECH (Uruguay).

Gráfico 6. Relación entre el stock de migrantes del conjunto de datos de Facebook y las estimaciones de stock de Naciones Unidas (UN) según país de origen y destino por sexo. Países seleccionados, 2020



Fuente: elaboración propia a partir de cifras de usuarios activos diarios de Facebook e *International Migrant Stock 2020* (Naciones Unidas).

Los resultados de los modelos multivariados muestran un importante nivel de correlación entre Facebook y los datos empleados como comparación, siendo mayor para el caso de Naciones Unidas (Tabla 4). Este resultado es esperable si se considera que, como fue mencionado en la sección de “Datos y metodología”, las estimaciones de Naciones Unidas de 2020 para buena parte de los países de destino se encuentran corregidas a partir de las cifras de solicitantes de asilo, personas bajo la condición de refugiado o desplazados venezolanos, lo cual tiende a incrementar el valor del stock derivado de la proyección demográfica.

Respecto a los sesgos estimados para los países origen, se observa que los coeficientes que son estadísticamente significativos al 95% de confianza tienen signo negativo, lo que implica que nuevamente Facebook sobreestima la magnitud del stock respecto al estándar de referencia considerado, en este caso las estimaciones de Naciones Unidas (i.e. Brasil, El Salvador, Haití, Honduras, Perú y República Dominicana) (Tabla 4). En lo que concierne a los contextos de destino, la significancia es variable según fuente de información, pero cabe

destacar el caso de Costa Rica, el cual solo cuenta con un origen de análisis que es Nicaragua. En particular, para 2022 (EH) y 2020 (UN), se observa una subestimación del stock de personas migrantes de origen nicaragüense por parte de las cifras de audiencias de Facebook (Tabla 5). Una de las posibles explicaciones que pueden estar por detrás de dicho desajuste tiene que ver con el bajo nivel de penetración de internet en los hogares de Nicaragua (Gráfico 8 en Anexo), que repercute en el número de usuarios de Facebook (Gráfico 7 en Anexo) y, por ende, tiene incidencia en el uso de la red social en el país de destino. Complementariamente y anidado con lo anterior, la subrepresentación de los datos de Facebook para dicho origen puede estar respondiendo a la presencia de una estructura etaria más envejecida de las personas que residen en Costa Rica que es consecuente con una tradición migratoria histórica (Castro Valverde, 2007; Gatica López, 2007; Baumeister, 2006).

En cuanto al regresor que alude al sexo, los resultados son coincidentes con lo encontrado por Varona (2022) y Prieto Rosas et al. (2022b), y muestran que respecto a ambas fuentes de información convencionales no existen diferencias significativas cuando se controla por el resto de las variables independientes que especifican el modelo (Tabla 4).

Tabla 4. Coeficientes del modelo de regresión lineal del logaritmo del stock de migrantes absolutos de origen i en el destino j y sexo k con base en las encuestas de hogares y Naciones Unidas (UN). Países seleccionados, 2020-2022

| | 2021 | | | | 2022 | | | | 2020 | | | |
|-------------------------|-----------------------------------|---------------------|----------------------|----------------------|-----------------------------------|----------------------|----------------------|----------------------|-------------------------|---------------------|---------------------|---------------------|
| | log(stock migrantes absolutos EH) | | | | log(stock migrantes absolutos EH) | | | | log(stock migrantes UN) | | | |
| | (1) | (2) | (3) | (4) | (1) | (2) | (3) | (4) | (1) | (2) | (3) | (4) |
| log(<i>expats</i> FBK) | 1.573*** (0.179) | 1.368*** (0.212) | 1.140*** (0.254) | 1.152*** (0.251) | 1.688*** (0.135) | 1.544*** (0.15) | 1.220*** (0.167) | 1.221*** (0.167) | 1.391*** (0.198) | 1.316*** (0.211) | 0.995** (0.302) | 0.998** (0.301) |
| País de destino | | | | | | | | | | | | |
| Argentina | | -1.003 (0.688) | -1.07 (0.703) | -1.065 (0.711) | | -0.37 (0.562) | -0.537 (0.568) | -0.536 (0.57) | | -0.217 (0.433) | -0.391 (0.464) | -0.389 (0.465) |
| Chile | | 1.234** (0.368) | 0.905 (0.492) | 0.901 (0.491) | | 0.923*** (0.263) | 0.409 (0.374) | 0.409 (0.375) | | 2.007*** (0.256) | 1.967*** (0.338) | 1.967*** (0.338) |
| Colombia | | 0.217 (0.421) | -0.469 (0.527) | -0.465 (0.543) | | 0.0141 (0.311) | -0.236 (0.419) | -0.235 (0.42) | | 0.580* (0.289) | 0.229 (0.401) | 0.231 (0.401) |
| Costa Rica | | 1.520** (0.477) | 1.789 (1.314) | 1.749 (1.312) | | 1.001** (0.347) | 3.467** (1.139) | 3.463** (1.143) | | 1.873*** (0.41) | 3.073** (1.053) | 3.065** (1.051) |
| Ecuador | | -0.582 (0.564) | -0.844 (0.635) | -0.831 (0.631) | | -0.422 (0.432) | -0.783 (0.481) | -0.782 (0.483) | | 0.407 (0.402) | 0.0622 (0.464) | 0.0649 (0.465) |
| Paraguay | | 1.049** (0.365) | 0.589 (0.67) | 0.596 (0.677) | | 0.54 (0.518) | 0.11 (0.7) | 0.111 (0.702) | | 1.460*** (0.323) | 0.749 (0.615) | 0.75 (0.618) |
| Perú | | -2.355*** (0.68) | -2.553*** (0.707) | -2.541*** (0.704) | | -2.202*** (0.557) | -2.493*** (0.585) | -2.491*** (0.587) | | 0.613 (0.314) | 0.223 (0.328) | 0.226 (0.328) |
| Uruguay | | 1.103* (0.482) | 0.139 (0.784) | 0.165 (0.792) | | 1.226*** (0.335) | 0.188 (0.528) | 0.191 (0.529) | | 1.063 (0.582) | -0.142 (0.851) | -0.137 (0.852) |
| País de origen | | | | | | | | | | | | |
| Argentina | | | -0.825 (0.557) | -0.806 (0.555) | | | -0.548 (0.426) | -0.546 (0.427) | | | -1.071 (0.588) | -1.067 (0.588) |
| Brasil | | | -1.47 (0.769) | -1.45 (0.76) | | | -1.105* (0.506) | -1.103* (0.505) | | | -1.378* (0.615) | -1.374* (0.614) |
| Chile | | | -0.769 (0.813) | -0.741 (0.812) | | | -0.897 (0.607) | -0.893 (0.606) | | | -0.644 (0.867) | -0.638 (0.866) |
| Colombia | | | -0.404 (0.54) | -0.391 (0.538) | | | -0.313 (0.426) | -0.312 (0.427) | | | -1.288* (0.5) | -1.286* (0.501) |
| Cuba | | | -1.515 (1.07) | -1.483 (1.067) | | | -0.245 (0.58) | -0.241 (0.582) | | | -1.622 (0.853) | -1.615 (0.852) |
| El Salvador | | | -4.852*** (1.311) | -4.820*** (1.304) | | | -2.804** (1.033) | -2.801** (1.037) | | | -3.074** (0.982) | -3.067** (0.982) |

| | | | | | | | | | | | | |
|-------------------------|----------------------|--------------------|--------------------|--------------------|----------------------|----------------------|---------------------|---------------------|---------------------|--------------------|---------------------|--------------------|
| Guatemala | | | -1.315 (1.302) | -1.277 (1.317) | | | -2.247* (0.924) | -2.242* (0.923) | | | -2.182 (1.107) | -2.174 (1.106) |
| Honduras | | | -3.073* (1.384) | -3.034* (1.374) | | | -2.924** (0.898) | -2.920** (0.907) | | | -2.336* (1.131) | -2.328* (1.129) |
| Haití | | | -0.868 (0.993) | -0.82 (0.969) | | | -0.396 (0.691) | -0.391 (0.689) | | | -2.128* (0.903) | -2.118* (0.904) |
| Nicaragua | | | -1.374 (1.447) | -1.325 (1.445) | | | -3.165** (1.102) | -3.160** (1.108) | | | -2.36 (1.215) | -2.35 (1.213) |
| Perú | | | -0.923 (0.638) | -0.902 (0.641) | | | -0.545 (0.397) | -0.543 (0.397) | | | -1.567* (0.692) | -1.563* (0.694) |
| R. Dominicana | | | -1.841 (1.037) | -1.807 (1.032) | | | -0.981 (0.838) | -0.977 (0.843) | | | -2.560** (0.979) | -2.553* (0.978) |
| Sexo | | | | 0.247 (0.337) | | | | 0.0225 (0.266) | | | | 0.0591 (0.186) |
| Constante | -6.883*** (1.715) | -4.665* (1.998) | -1.044 (3.017) | -1.554 (2.984) | -8.001*** (1.321) | -6.432*** (1.452) | -2.082 (1.982) | -2.131 (2.03) | -4.967** (1.819) | -4.854* (2.013) | -0.0957 (3.576) | -0.214 (3.542) |
| N | 140 | 140 | 140 | 140 | 154 | 154 | 154 | 154 | 154 | 154 | 154 | 154 |
| R ² | 0.528 | 0.632 | 0.717 | 0.718 | 0.616 | 0.703 | 0.775 | 0.775 | 0.694 | 0.756 | 0.815 | 0.816 |
| R ² ajustado | 0.524 | 0.606 | 0.666 | 0.665 | 0.614 | 0.685 | 0.739 | 0.737 | 0.692 | 0.74 | 0.786 | 0.785 |

Nota: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Error estándar robusto a heteroscedasticidad entre paréntesis. Se empleó *dummy* y *contrast coding*, y para evitar el problema de especificación de colinealidad perfecta se omitió a México como destino y Venezuela como origen.

Fuente: elaboración propia a partir de cifras de usuarios activos diarios de Facebook, microdatos de la EPH (Argentina), ENE (Chile), GEIH (Colombia), ENH (Costa Rica), ENEMDU (Ecuador), ENOE (México), EPH (Paraguay), ENAHO (Perú) y ECH (Uruguay) y estimaciones de *International Migrant Stock 2020* (Naciones Unidas).

Conclusiones

Los desafíos que enfrenta el estudio de la migración internacional en América Latina se han incrementado considerablemente producto de las transformaciones más recientes sufridas en torno a la magnitud, dirección, composición demográfica y grado de voluntariedad del movimiento. En la última década ha quedado claro que los censos de población no son suficientes para captar el dinamismo de un sistema migratorio en transformación (Bengochea y Pellegrino, 2023; Prieto Rosas y Zapata, 2023; Prieto Rosas y Bengochea, 2022).

En paralelo, las Huellas Digitales como fuente de información relevante para el estudio de la migración internacional ha ido ganado terreno, primero en la investigación demográfica de países del norte global (Spyratos et al., 2019, 2020; Zagheni, Weber y Gummadi, 2017), y más recientemente, en el contexto latinoamericano. En el segundo caso, comienza a acumularse la evidencia que encuentra una importante correlación de las Huellas Digitales derivadas del uso de Facebook con fuentes de información “tradicional” como registros administrativos (Palotti et al., 2020), estimaciones de stock de migrantes de Naciones Unidas (Prieto Rosas et al., 2022b), datos censales (Varona, 2022) y encuestas de hogares (Rosati, Cerruti y Maguire, 2021).

En este escenario de desafíos metodológicos, el presente trabajo de tesis tuvo como propósito contribuir al acervo de investigación descrito y buscó analizar en qué medida los datos de usuarios de Facebook etiquetados como “anteriormente viviendo en...” se aproximan a la cuantificación de poblaciones migrantes que arrojan las encuestas de hogares en un ejercicio sistemático para varios países latinoamericanos en 2021 y 2022. Para ello, se tomó como principal referencia la metodología empleada por Zagheni, Weber y Gummadi (2017), y se estimaron una serie de modelos lineales multivariados que brindaron información sobre el nivel de correlación entre Huellas Digitales del uso de Facebook y las encuestas de hogares de nueve países latinoamericanos, así como sobre la dirección y magnitud de los sesgos por país de origen, destino y sexo. La elección de las encuestas de hogares como estándar de comparación respondió a la ventaja relativa que tiene esta fuente derivada de su periodización de recolección anual.

Los principales resultados del método de calibración con base a encuestas de hogares pueden resumirse en cuatro grandes líneas. Primero, la identificación de una alta correlación entre el DAU de *expats* de Facebook y los datos de stocks de migrantes derivados de encuestas de hogares que es variable según el país de destino y origen considerado. De este modo, es posible sostener que las cifras de *expats* de Facebook representan una fuente de información

complementaria de relevancia para la medición de la migración internacional en la región, y adquiere un valor superlativo en tanto permite monitorear en tiempo real las variaciones en las magnitudes de los stocks poblacionales. Incluso, esta idea se ve reforzada si se compara el nivel de correlación encontrado entre Facebook y las encuestas de hogar, con el de Facebook y Naciones Unidas, especialmente para el año 2022.

Segundo, se encuentra una tendencia general a la sobrerrepresentación de los datos digitales respecto a las dos fuentes de información convencionales seleccionadas como referencia –encuestas de hogares y Naciones Unidas–. Este resultado es opuesto a lo encontrado por Zagheni, Weber y Gumradi (2017) para un contexto de destino del norte global como Estados Unidos, quienes identifican una sistemática subrepresentación de los stocks de migrantes por parte de Facebook respecto a la *American Community Survey*. En tanto, Varona (2022) halla que para el caso mexicano el ajuste en la estimación entre fuentes es variable según país de nacimiento, pero en general hay un patrón compartido a la subrepresentación de las magnitudes en los datos de Facebook. En cambio, los resultados de este trabajo coinciden con los alcanzados por Rosati, Cerruti y Maguire (2021) para nacionales venezolanos a partir de la Encuesta Anual de Hogares para la Ciudad Autónoma de Buenos Aires y por Spyrtos et al. (2019) para las personas migrantes latinoamericanas en la *American Community Survey*. Por otra parte, los modelos estimados para nueve países de la región muestran que la correlación entre Facebook y Naciones Unidas es mayor a la identificada para el binomio Facebook-encuestas de hogares cuando se considera a los migrantes absolutos según país de nacimiento.

Los resultados destacados pueden estar respondiendo a la conjugación de diversos factores vinculados a la naturaleza de las tres fuentes consideradas en el análisis. Es esperable que los stocks de migrantes estimados a partir de las encuestas de hogares sean menores a las magnitudes reportadas por Facebook porque son resultado del tratamiento con muestras, y como se ha mencionado a lo largo del documento, la imprecisión de la cuantificación de poblaciones migrantes según origen nacional es un aspecto a considerar. También, es previsible que el ajuste respecto a Naciones Unidas sea mayor ya que las estimaciones de *International Migrant Stock* incluyen cifras de registros administrativos, lo cual presiona al alza las magnitudes de los stocks por díadas origen-destino, aunque no es conveniente desestimar la corrección imputada por efecto de la estructura etaria. Complementariamente, se ha documentado que la cuantificación del DAU y MAU de Facebook está “inflada” por la presencia de usuarios con más de una cuenta o *bots*, y la prevalencia de esta limitación es mayor en países en desarrollo, como es el caso de aquellos de la región latinoamericana, respecto a los contextos desarrollados (Rampazzo et al., 2021; U.S. SEC, 2019). Este aspecto podría

explicar, en parte, la dirección opuesta de los resultados encontrados respecto a Zagheni, Weber y Gummadi (2017). Sin embargo, aún queda pendiente comprender si la incidencia de cuentas múltiples o falsas difiere según origen y destino y/o entre poblaciones migrantes y generales, aunque es sensato pensar que las personas en situación de movilidad tienen pocos incentivos para cambiar de usuario dado el uso que le dan a las redes sociales web como espacio de búsqueda de información durante el viaje y la llegada (Rampazzo et al., 2021; Gandini, Fernández de la Reguera y Narváez Gutiérrez, 2020). Entonces, ante la incógnita de si lo que pesa más en los resultados encontrados es la subenumeración de las encuestas o la sobreestimación de los datos Facebook, y considerando los resultados que encuentra Varona (2022) a partir de información censal, podría pensarse que la sobrerrepresentación identificada es fruto fundamentalmente del diseño muestral de las encuestas (i.e. exclusión de viviendas colectivas, desactualización del marco muestral, expansores de muestra contruidos con base a proyecciones poblacionales desactualizadas). Validar esta hipótesis, queda como una línea de indagación futura que podría aprovechar próximamente los resultados de la ronda censal 2020 de la región.

Tercero, los modelos lineales multivariados muestran un importante nivel de correlación para las dos definiciones consideradas, migrantes absolutos y recientes según país de nacimiento, aunque se identifica un ajuste mayor para el primer grupo. Este resultado cobra sentido si se parte de la base de que ya se identifica un comportamiento general a la sobrerrepresentación de los datos de Facebook respecto a las encuestas de hogares para los migrantes absolutos, y considerando que los recientes son un subgrupo de éstos, la brecha entre fuentes será necesariamente más amplia. Asimismo, si se compara lo encontrado por Varona (2022) para las mismas definiciones, los resultados son coincidentes. De este modo, es posible afirmar que, incluso para países con historias migratorias bien distintas, como es el caso de Argentina –receptor de larga data– y Perú –de recepción neta más reciente–, Facebook es una fuente complementaria relevante para obtener estimaciones más precisas de la migración internacional en la región. Sin embargo, como ya fue mencionado y se retoma inmediatamente, ninguna de las definiciones a partir de las cuales se ha evaluado el ajuste de Facebook respecto a una fuente de información tradicional es estrictamente comparable con la segmentación de públicos según “vivieron anteriormente en...”.

Cuarto, en concordancia con lo encontrado por Prieto Rosas et al. (2022b), Varona (2022) y Rosati, Cerruti y Maguire (2021) los modelos estimados que incluyen al sexo como regresor muestran que las diferencias entre fuentes de información no son significativas por sexo una vez que se controla por el país de destino, origen y el logaritmo del número de *expats*.

Consecuentemente, parece no ser necesario incluir el sexo como variable de control en futuras estimaciones, lo cual es una decisión tomada inicialmente por Zagheni, Weber y Gummadi (2017).

Complementariamente, es relevante destacar dos resultados adicionales. Por un lado, el ajuste entre fuentes de información parece ser más importante para los orígenes nacionales de mayor relevancia en términos numéricos de cada país de destino, de manera tal que, la proximidad en las estimaciones para los nacionales venezolanos es una constante entre los contextos de destino analizados. Por otro lado, tal como ya había sido reconocido por Gutiérrez et al. (2020), y se observa en el análisis gráfico, las encuestas de hogares presentan importantes limitaciones cuando se quiere trabajar con orígenes nacionales de baja representación, lo cual lleva a plantear la idea de que las Huellas Digitales son una fuente a tener en cuenta para medir la migración internacional cuando se trate de orígenes poco numerosos en el contexto de destino.

Limitaciones y futuro trabajo

Vale la pena destacar una serie de limitaciones que enfrenta el análisis expuesto y que inciden sobre los resultados alcanzados y las reflexiones finales anteriormente desarrolladas. Primero, la naturaleza de los datos derivados de Huellas Digitales de Facebook enfrenta dos restricciones: la imposibilidad de exclusión de cuentas que pertenezcan a *bots* o a un mismo usuario según díadas origen-destino y el desconocimiento de los criterios y metodología empleada para la construcción de la segmentación del público según lugar de residencia actual y anterior.

Segundo, se presentan restricciones para la estimación de manera directa de sesgos robustos y representativos por origen para cada país de destino que sirvan de factor de corrección para futuras estimaciones en tiempo real de stocks de migrantes con cifras de usuarios “anteriormente viviendo en...” de Facebook. No obstante, las estimaciones provistas podrían actualizarse anualmente para obtener correctores de sesgos para el conjunto de orígenes de cada país de destino o bien para cada origen sin desagregación por país de residencia actual. Esta es una clara diferencia con los correctores estimados a partir de información censal o estimaciones de Naciones Unidas, dado que la actualización de las cifras de stocks se encuentra más dilatada en el tiempo y por consiguiente no es posible hacer una reestimación anual.

Tercero, si bien se ha documentado para el sistema migratorio latinoamericano que la propensión a migrar es disímil entre etapas del ciclo de vida (Prieto Rosas, 2013) y el uso de las redes sociales web también se encuentra segmentado por la edad (Gil-Clavel, Zagheni y Bordone, 2022), no fue posible incluir a dicha variable demográfica como regresor de control de la estimación de los sesgos entre fuentes producto de las limitaciones que imponen el tamaño muestral de las encuestas de hogares para la desagregación de las cifras de stocks de población migrante.

Cuarto, vale la pena señalar que la lista de países de residencia anterior para la que Facebook proporciona cifras de usuarios no es exhaustiva, quedando por fuera orígenes de relevancia para ciertos contextos de destino latinoamericanos, como es el caso de las personas de origen paraguayo y boliviano en Argentina. Este aspecto representa una limitación de relevancia cuando se busca tener una aproximación cabal del escenario migratorio de ciertos países de destino de la región.

Quinto, ninguna de las definiciones de migrantes hasta ahora construidas en censos y encuestas es estrictamente comparable con la etiqueta “anteriormente viviendo en...” ya que no se recupera el lugar de residencia previo, sino el que alude a un período determinado anterior, generalmente delimitado por los cinco años previos. Sin embargo, los resultados muestran que aún, considerando definiciones *proxy* de la etiqueta proporcionada por Facebook, la correlación respecto a fuentes de información convencionales es elevada.

Sexto, una de las limitaciones inherentes a cualquier análisis que aborde la movilidad internacional, tiene que ver con que ninguna de las fuentes de información más tradicionales, incluso aquellas con periodicidad de recolección mayor a la de los censos, está exenta de debilidades, siendo difícil establecer un “estándar de oro” contra el que evaluar la calidad de los datos de Facebook u otros tipos de datos semejantes. Ante ello, es preciso explicitar los límites que presenta cada fuente empleada como referencia en la medida que afectan los resultados de la comparación establecida con las Huellas Digitales, tanto como los de los propios datos de Facebook. Complementariamente, parte de la literatura parece mostrar que las ventajas de implementación del paradigma bayesiano son mayores sobre el abordaje frecuentista para la medición de la migración internacional (Bengochea, 2022). El primero de estos abordajes, permite considerar múltiples fuentes de información, jerarquizarlas e informar modelos con criterios teóricos que tengan en cuenta información específica sobre los flujos, efectos período, e información emergente sobre los datos digitales (Rampazzo et al., 2021; Alexander, Polimis y Zagheni, 2020).

Como futuras líneas de investigación relevantes en materia del uso de Huellas Digitales, y en particular de Facebook, para el estudio de la movilidad internacional en América Latina se plantean las siguientes dos. En primer lugar y como ya se ha adelantado, la ronda censal 2020 presenta una oportunidad de especial importancia para continuar con el análisis de validez de los datos de Facebook, dada su pretensión de universalidad y la posibilidad de realizar comparaciones segmentando de manera simultánea a las poblaciones según origen, sexo y edad. Asimismo, es una excelente oportunidad para continuar profundizando en la estimación de los sesgos según definición de migrante adoptada, y formularios como el de Brasil y Uruguay brindan la posibilidad de identificar a las personas migrantes según lugar de residencia anterior, lo que abre la ventana de construcción de una categoría estrictamente comparable con la definición proporcionada por Facebook. De momento, dado que son muy escasos los países de la región que han realizado la operación censal y menos los que han publicado los microdatos, solo se ha podido realizar el ejercicio de cotejo para el censo de población de México (Varona, 2022). También, se abre la posibilidad de realizar el análisis de comparación entre fuentes a nivel subnacional, lo que representaría una novedosa contribución a la literatura sobre la temática que tiene como antecedente el trabajo realizado por Rosati, Cerruti y Maguire (2021). En segundo lugar, se considera de relevancia indagar en qué medida es posible identificar a los orígenes no contemplados dentro de las clasificaciones de usuarios “anteriormente viviendo en...” disponibles, a partir de la combinación de otras variables de segmentación de audiencias que proporciona Facebook Ads Manager, tales como la clasificación de intereses.

Referencias bibliográficas

- Alburez-Gutierrez, D., Zagheni, E., Aref, S., Gil-Clavel, S., Grow, A., & Negraia, D. V. (2019). Demography in the digital era: New data sources for population research. En G. Arbia, S. Peluso, A. Pini & G. Rivellini (Eds.), *Smart statistics for smart applications: book of short papers SIS2019* (pp. 23–30), Pearson.
- 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.
- Baumeister, E. (2006). Migración internacional y desarrollo en Nicaragua. *Serie Población y Desarrollo*, CEPAL, no. 67.
- Bengochea, J. (2014). Inmigración reciente en Uruguay: 2005 – 2011 (Tesis de Maestría, Programa de Población – Universidad de la República). Recuperado de <https://www.colibri.udelar.edu.uy/jspui/handle/20.500.12008/8151>
- Bengochea, J. (2018). Los movimientos migratorios de población sur-sur en América Latina: características del sistema migratorio y factores asociados a la migración, 1960-2010 (Tesis de Doctorado, El Colegio de México). Recuperado de <https://repositorio.colmex.mx/concern/theses/kp78gg509?locale=es>
- Bengochea, J. (2022). Huellas digitales y oportunidades para el estudio de la migración en América Latina y el Caribe. *Documento de Trabajo del Programa de Población*, no. 8.
- Bengochea, J., Cabezas, G., Gandini, L., Herrera, G., ..., & Zapata, G. P. (2022). COVID-19 y población migrante y refugiada. Análisis de las respuestas político institucionales en ciudades receptoras de seis países de América Latina. *Documento de Trabajo de CAMINAR*, no. 5.
- Bengochea, J., Del Fava, E., Prieto Rosas, V., & Zagheni, E. (2021). Leveraging census data to study migration flows in Latin America and the Caribbean: an assessment of the available data sources. *MPIDR Working Papers*, no. 19.
- Bengochea, J., & Madeiro, V. (2020). Acceso a la vivienda adecuada de las personas migrantes en la ciudad de Montevideo. Recuperado de https://omif.uy/wp-content/uploads/2020/11/INFORME_VIVIENDA_WEB.pdf
- Bengochea, J., & Pellegrino, A. (2023). Etapas de la migración internacional e intrarregional en América Latina y el Caribe. *Notas de Población*, 50(116), 137-157.
- Billari, F., D'Amuri, F., & Marcucci, J. (2016). Forecasting births using Google. En *Carma 2016: 1st international conference on advanced research methods in analytics* (pp. 119-119), Universitat Politècnica de València.
- Billari, F. C., & Zagheni, E. (2017). Big data and population processes: A revolution? En A. Petrucci y R. Verde (Eds.), *Statistics and Data Science: new challenges, new generations* (pp. 167–178), Firenze University Press.
- 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.

- Canales Cerón, A. I., & Rojas Wiesner, M. L. (2018). Panorama de la migración internacional en México y Centroamérica. *Serie Población y Desarrollo*, CEPAL, no. 124.
- Cárdenas Castro, M., & Yáñez Yáñez, S. (2012). Nuevas formas de muestreo para minorías y poblaciones ocultas: muestras por encuestado conducido en una población de inmigrantes sudamericanos. *Universitas Psychologica*, 11(2), 571-578.
- Carrasco, J. I., & Baizán, P. (2023). Integración al mercado laboral de los inmigrantes intrarregionales en Chile: análisis de las trayectorias de empleo y de los logros ocupacionales basado en las encuestas de hogares de 2013, 2015 y 2017. *Notas de Población*, 50(116), 159-193.
- Carrasco, I., & Suárez, J. I. (2018). Migración internacional e inclusión en América Latina: Análisis en los países de destino mediante encuestas de hogares. *Serie Políticas Sociales*, CEPAL, no. 231.
- Castro Valverde, C. (2007). Dimensión cuantitativa de la inmigración nicaragüense en Costa Rica: del mito a la realidad. En C. Sandoval (Ed.), *El mito roto. Inmigración y emigración en Costa Rica* (pp. 25-50), Editorial Universidad de Costa Rica.
- CEPAL. (2022). Seminario sobre los efectos de la pandemia en las encuestas de hogares. Recuperado de <https://www.cepal.org/es/eventos/seminario-efectos-la-pandemia-encuestas-hogares>
- Cerrutti, M. (2009). Gender and Intra- Regional Migration in South America. *Human Development Research Paper UNDP*, no. 12.
- Cerruti, M. (2022). COVID 19 y su impacto en la migración y la distribución de la población. *Cuarta reunión de la Conferencia Regional sobre Población y Desarrollo de América Latina y el Caribe*, CEPAL.
- Cerrutti, M., Fargues, P., & Awumbila, M. (2021). The Case for a World Migration Survey. *IUSSP - Policy & Research Papers*, no. 25.
- Cerruti, M., & Parrado, E. (2015). Migración intrarregional en América del Sur: tendencias y una agenda de investigación. *Revista Anual de Sociología*, 41, 399-421.
- 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
- Del Fava, E., Wiśniowski, A., & Zagheni, E. (2019). Modelling international migration flows by integrating multiple data sources. *SocArXiv Papers*. <https://doi.org/10.31235/osf.io/cma5h>
- Donzowa, J., Perrotta, D., & Zagheni, E. (2023). Assessing self-selection biases in online surveys: Evidence from the COVID-19 Health Behavior Survey. *MPIDR Working Paper*, (47).
- Edwards, R., & Greene, M. (2022). Migración haitiana en Santiago. Una aproximación multiescalar y temporal. *EURE (Santiago)*, 48(144), 1-21.
- Efron, B., & Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*, Chapman and Hall/CRC.
- Fiorio, L., Abel, G., Cai, J., Zagheni, E., Weber, I., & Vinué, G. (2017). Using twitter data to estimate the relationship between short-term mobility and long-term migration. En *Proceedings of the 2017 ACM on web science conference* (pp. 103-110).
- Gandini, L., de la Reguera, A. F., & Narváez Gutiérrez, J. C. (2020). *Caravanas*, UNAM-Secretaría de Desarrollo Institucional.

- Garcés Mascareñas, B., & López-Sala, A. (2021). Inmóviles, varados y excluidos: los efectos de la COVID-19 en el régimen internacional de asilo. *Revista CIDOB d'afers internacionals*, (129), 7-29.
- Gatica López, G. (2007). Migración nicaragüense a Costa Rica y políticas públicas. En C. Sandoval (Ed.), *El mito roto. Inmigración y emigración en Costa Rica* (pp. 113-144), Editorial Universidad de Costa Rica.
- Gendronneau, C., Wiśniowski, A., Yildiz, D., Zagheni, E., Fiorio, L., Hsiao, Y., ..., & Hoorens, S. (2019). Measuring labour mobility and migration using big data: exploring the potential of social-media data for measuring EU mobility flows and stocks of EU movers. Recuperado de <https://op.europa.eu/en/publication-detail/-/publication/6c4b29f7-6e47-11ea-b735-01aa75ed71a1/language-en>
- Gil-Clavel, S., & Zagheni, E. (2019). Demographic differentials in Facebook usage around the world. *Proceedings of the International AAAI Conference on Web and Social Media*, 13(1), 647–650.
- Gil-Clavel, S., Zagheni, E., & Bordone, V. (2022). Close Social Networks Among Older Adults: The Online and Offline Perspectives. *Population Research and Policy Review*, 41(3), 1111–1135.
- 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.
- Giorguli, S., Lindstrom, D., Nájera, J., Prieto, V., Márquez, C., & Amaro, M. (2023). Plataforma de Datos Territoriales para la Integración de Inmigrantes. Etnoencuesta de Inmigración Reciente en Contextos de Acogida Latinoamericanos LAMP-ENIR 2021. Recuperado de <https://mmp-lamp.colmex.mx/wp-content/uploads/informe-de-resultados-lamp-emir-2021.pdf>
- 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.
- Hargittai E. (2007). Whose space? Differences among users and non-users of social network sites. *Journal of Computer-Mediated Communication*, 13, 276–297. <https://doi.org/10.1111/j.1083-6101.2007.00396.x>
- Hargittai E. (2015). Is bigger always better? Potential biases of big data derived from social network sites. *The ANNALS of the American Academy of Political and Social Science*, 659, 63–76. <https://doi.org/10.1177/0002716215570866>
- Hargittai, E. (2018). Potential Biases in Big Data: Omitted Voices on Social Media, *Social Science Computer Review*, 38(1), 10–24. <https://doi.org/10.1177/0894439318788322>
- Herdağdelen, A., State, B., Adamic, L., & Mason, W. (2016). The social ties of immigrant communities in the United States. En *Proceedings of the 8th ACM Conference on Web Science* (pp. 78-84).

- Herrera, G., & Gómez, C. (2022). Introduction: Emergent Issues of South American Migrations. En G. Herrera y C. Gómez (Eds.), *Migration in South America* (pp. 1-26), IMISCOE.
- Hitsch, G. J., Hortaçsu, A., & Ariely, D. (2010). What makes you click? Mate preferences in online dating. *Quantitative marketing and Economics*, 8, 393-427.
- Hughes, C., Zagheni, E., Abel, G. J., Sorichetta, A., Wi'sniowski, A., Weber, I., & Tatem, A. J. (2016). Inferring migrations: traditional methods and new approaches based on mobile phone, social media, and other big data: feasibility study on inferring (labour) mobility and migration in the European union from big data and social media data. Recuperado de https://www.researchgate.net/profile/Ingmar-Weber/publication/305984089_Inferring_Migrations_Traditional_Methods_and_New_Approaches_based_on_Mobile_Phone_Social_Media_and_other_Big_Data/links/57a816e508ae455e854703fc/Inferring-Migrations-Traditional-Methods-and-New-Approaches-based-on-Mobile-Phone-Social-Media-and-other-Big-Data.pdf
- 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.
- Kashyap, R. (2021). Has demography witnessed a data revolution? Promises and pitfalls of a changing data ecosystem. *Population Studies*, 75(1), 47-75.
- Kashyap, R., Rinderknecht, R. G., Akbaritabar, A., Albrez-Gutierrez, D., Gil-Clavel, S., Grow, A., ..., & Zhao, X. (2022). Digital and Computational Demography. *SocArXiv Papers*. <https://doi.org/10.31235/osf.io/7bvpt>
- Koolhaas, M. (2023). Introducción a la migración. En Prieto, V. y Robello, M. (Coord.) *Manual de Demografía* (pp. 163-178), Programa de Población-Udelar.
- Latinobarómetro. (2020). Datos. Recuperado de <https://www.latinobarometro.org/latContents.jsp>
- Livi Bacci, M. (1993). *Introducción a la demografía*, Ariel.
- McAuliffe, M., Freier, L. F., Skeldon, R., & Blower, J. (2021). El gran perturbador: Impacto mundial de la COVID-19 en la migración, la movilidad y los migrantes. En M. McAuliffe y A. Triandafyllidou, (Eds.) *Informe sobre las Migraciones en el Mundo 2022*, Organización Internacional para las Migraciones (OIM).
- Maguid, A. M. (2008). La emigración internacional a través de los censos en países de origen: evaluación de resultados y recomendaciones. *Serie Población y Desarrollo*, CEPAL, no. 86.
- Márquez, C., Prieto Rosas, V., & Escoto, A. (2020). Segmentación en el ingreso por trabajo según condición migratoria, género y ascendencia étnico-racial en Uruguay. *Revista Migraciones*, (48), 85-118.
- Markey, P. M., & Markey, C. N. (2013). Seasonal variation in internet keyword searches: a proxy assessment of sex mating behaviors. *Archives of sexual behavior*, 42, 515-521.
- Martínez Pizarro, J. (2005). Magnitud y dinámica de la inmigración en Chile, según el Censo de 2002. *Papeles de población*, 11(44), 109-147.
- 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., Cano Christiny, M. V., & Soffia Contrucci, M. (2014). Tendencias y patrones de la migración latinoamericana y caribeña hacia 2010 y desafíos para una agenda regional. *Serie Población y Desarrollo*, CEPAL, no. 109.
- 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.
- Masferrer, C., & Prieto, V. (2019). El perfil sociodemográfico del retorno migratorio reciente. Diferencias y similitudes entre contextos de procedencias y de acogida en América Latina. En L. Rivera-Sánchez (Ed.), *¿Volver a casa? Migrantes de retorno en América Latina. Debates, tendencias y experiencias divergentes* (pp. 67-126), Colegio de México.
- Melella, C. E., & Perret, G. (2016). El uso de Internet en contextos migratorios: Una aproximación a su estudio. *Enl@ce: Revista Venezolana de Información, Tecnología y Conocimiento*, 13(2).
- Morales Gamboa, A. (2008). Inmigración en Costa Rica: características sociales y laborales, integración y políticas públicas. *Serie Población y desarrollo*, (85).
- Moran, C. (2023). The ‘connected migrant’: A scoping review. *Convergence*, 29(2), 288-307.
- Naciones Unidas. (2020a). International Migrant Stock 2020. Recuperado de <https://www.un.org/development/desa/pd/content/international-migrant-stock>
- Naciones Unidas. (2022b). Methodology Report. International Migrant Stock 2020. Recuperado de https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/undesapd_2020_international_migrant_stock_documentation.pdf
- Nájera, J. N. (2016). El complejo estudio de la actual migración en tránsito por México: Actores, temáticas y circunstancias. *Migraciones internacionales*, 8(3), 255-266.
- Nájera, J. N. (2017). Migración, fuerza de trabajo y familia, elementos en la definición del espacio transfronterizo México-Guatemala. *EntreDiversidades*, (8), 119-150.
- OIM. (2021). Grandes movimientos de migrantes altamente vulnerables en las Américas provenientes del Caribe, Latinoamérica y otras regiones. Recuperado de https://rosanjose.iom.int/sites/g/files/tmzbd11446/files/documents/oim_grandes_movimientos_de_migrantes_altamente_vulnerables_en_las_americas_spavf_0.pdf
- OIM. (2022a). Tendencias recientes de la migración en las Américas. Recuperado de https://rosanjose.iom.int/sites/g/files/tmzbd11446/files/documents/tendencias-recientes-de-la-migracion-en-las-americas_sp.pdf
- OIM. (2022b). Migraciones sur-norte desde Sudamérica. Rutas, vulnerabilidades y contextos del tránsito de migrantes extrarregionales. Recuperado de https://robuenosaires.iom.int/sites/g/files/tmzbd1626/files/documents/oim_migraciones-sur-norte-desde-suramerica.pdf
- OIM. (2023). DTM South America. Recuperado de <https://dtm.iom.int/regions/south-america>
- Ojala, J., Zagheni, E., Billari, F., & Weber, I. (2017). Fertility and its meaning: Evidence from search behavior. En *Proceedings of the International AAAI Conference on Web and Social Media*, 11(1), 640-643.
- 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.
<https://doi.org/10.1371/JOURNAL.PONE.0229175>
- Pellegrino, A. (2009). Las migraciones entre los países del Mercosur: tendencias y características. En Observatorio de Políticas Públicas de Derechos Humanos en el Mercosur (Ed.), *Las migraciones humanas en el Mercosur. Una mirada desde los derechos humanos. Compilación normativa* (pp. 17-26), Observatorio de Políticas Públicas de Derechos Humanos en el Mercosur.
- 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*, (2).
- Pestre, G., Letouzé, E., & Zagheni, E. (2016). The ABCDE of Big Data: Assessing Biases in Call-detail records for Development Estimates. Recuperado de <https://thedocs.worldbank.org/en/doc/551311466182785065-0050022016/original/PestreLetouzeZagheniABCDEMay2016.pdf>
- Plataforma R4V. (2023). Agosto 2023. Refugiados y migrantes venezolanos en la región. Recuperado de <https://www.r4v.info/es/document/r4v-america-latina-y-el-caribe-refugiados-y-migrantes-venezolanos-en-la-region-ago-2023>
- PNUD. (2023). Cruzando fronteras: El crecimiento sin precedentes de la migración dentro de América Latina y el Caribe. Recuperado de <https://www.undp.org/es/latin-america/blog/cruzando-fronteras-el-crecimiento-sin-precedentes-de-la-migracion-dentro-de-america-latina-y-el-caribe>
- Preston, S., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and modeling population processes*, Blackwell Publishing.
- Prieto Rosas, V. (2015). El Componente demográfico en las migraciones exteriores de América Latina, 1950-2050 (Tesis de Doctorado, Universidad Autónoma de Barcelona). Recuperado de <https://ddd.uab.cat/record/114008>
- Prieto Rosas, V., & Bengochea, J. (2022). International migration in South America. En A. E. Feldmann, X. Bada, J. Durand & S. Schütze (Eds.), *The Routledge History of Modern Latin American Migration* (pp. 62–77), Taylor & Francis.
- Prieto Rosas, V., Bengochea, J., Fernández Soto, M., Márquez Scotti, C., & Montiel, C. (2022). Informe de resultados de la Etnoencuesta de Inmigración Reciente en Montevideo (ENIR 1, 2018). *Documento de Trabajo del Programa de Población*, no. 7.
- Prieto Rosas, V., Bengochea, J., Pedemonte, M., & Zagheni, E. (2022b). Using Facebook to estimate migrant stock in Latin American and Caribbean countries, *MPIDR Lab in Digital and Computational Demography Workshop*.
- Prieto Rosas, V., & Márquez Scotti, C. (2019). Inclusión social de inmigrantes recientes que residen en viviendas particulares de Uruguay. *Documento de Trabajo del Programa de Población*, no. 4.
- Prieto Rosas, V., & Montiel, C. (2020). Inclusión social de niños, niñas y adolescentes vinculados a la inmigración. Recuperado de https://omif.cienciassociales.edu.uy/wpcontent/uploads/2020/11/Inclusion-Social-inmigracionUruguay_web.pdf

- Prieto Rosas, V., Montiel, C., Barasch, G., & Bengochea, J. (2022a). Validating Facebook Tagging to Examine the Impacts of COVID-19 Pandemic on a South-South Migration Setting. *Conference of the Population Association of America*.
- Prieto Rosas, V., Montiel, C., Montoli, P., & Riaño, M. E. (en prensa). Informe de resultados. Encuesta sobre Migración e impactos de la pandemia por COVID19.
- Prieto Rosas, V. & Zapata, G. P. (2023). Unequal Origins to Unequal Destinations: Trends and Characteristics of Migrants' Social and Economic Inclusion in South America. En H. Crawley & J. Kofi Teye (Eds.) *The Palgrave Handbook of South–South Migration and Inequality* (pp. 247-269), Springer International Publishing.
- Proyecto COVID-19 e In(movilidad) en las Américas. (s.f.). Migración en reversa. Recuperado de <https://www.inmovilidadamericas.org/en-reversa>
- 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.
- Sánchez Aguilar, A. (2012). Perfil Migratorio del Perú. Recuperado de <https://publications.iom.int/books/perfil-migratorio-del-peru-2012>
- 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.
- Spyratos, S., Vespe, M., Natale, F., Weber, I., Zagheni, E., & Rango, M. (2018). Migration data using social media. Recuperado de https://publications.jrc.ec.europa.eu/repository/bitstream/JRC112310/facebook_study_v3.pdf
- Stefoni, C. (2018). Panorama de la migración internacional en América del Sur. Documento elaborado en el marco de la Reunión Regional Latinoamericana y Caribeña de Expertas y Expertos en Migración Internacional preparatoria del Pacto Mundial para una Migración Segura, Ordenada y Regular. *Serie Población y Desarrollo*, CEPAL, no. 123.
- Stern, M. J., Bilgen, I., McClain, C., & Hunscher, B. (2017). Effective sampling from social media sites and search engines for web surveys: Demographic and data quality differences in surveys of Google and Facebook users. *Social Science Computer Review*, 35, 713–732. <https://doi.org/10.1177/0894439316683344>

- Stewart, I., Flores, R. D., Riffe, T., Weber, I., & Zagheni, E. (2019). Rock, rap, or reggaeton?: assessing mexican immigrants' cultural assimilation using facebook data. En *The world wide web conference* (pp. 3258-3264).
- Tapinos, G. (1988). *Elementos de Demografía*, Espasa-Universidad.
- Tyldum, G., & Johnston, L. G. (2014). *Applying Respondent Driven Sampling to Migrant Populations. Lessons from the Field*, Palgrave Macmillan.
- UCLA. (s.f.). R library contrast coding systems for categorical variables. Recuperado de <https://stats.oarc.ucla.edu/r/library/r-library-contrast-coding-systems-for-categorical-variables/>
- U.S. SEC. (2019). Facebook, Inc.: 2018 annual report, form 10-K. Recuperado de <https://www.sec.gov/Archives/edgar/data/1326801/000132680119000009/fb-12312018x10k.htm>
- Varona, T. (2022). Análisis de la validez de datos de Facebook para el estudio de la inmigración en México (Tesis de Maestría, El Colegio de México). Recuperado de 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
- Vera, F., & Adler, V. (2020). *Inmigrando: Fortalecer ciudades destino. Tomo 1*, Banco Interamericano de Desarrollo.
- Vieira, C., Ribeiro, F., Vaz de Melo, P. O., Benevenuto, F., & Zagheni, E. (2020). Using Facebook data to measure cultural distance between countries: the case of Brazilian cuisine. En *Proceedings of the web conference 2020* (pp. 3091-3097).
- Wanner, P. (2021). How well can we estimate immigration trends using Google data? *Quality & Quantity*, 55, 1181–1202.
- 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>
- Willekens, F., Massey, D., Raymer, J., & Beauchemin, C. (2016). International migration under the microscope. *Science*, 352, 897–899.
- Wisniowski, A., Yıldız, D., Abel, G., Weber, I., Zagheni, E., & Hoorens, S. (2020). Measuring Migration Stocks Using Traditional and Social Media Data, *BigSurv20 Conference*.
- 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. *Proceedings of the 4th Annual ACM Web Science Conference, WebSci'*, 348–357. <https://doi.org/10.1145/2380718.2380764>
- Zagheni, E., & Weber, I. (2015). Demographic research with non-representative internet data. *International Journal of Manpower*, 36(1), 13–25.
- 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.

Anexo

Cuadro 8. Síntesis de la disponibilidad de información sobre migración internacional en las encuestas de hogares de América Latina y el Caribe, 2020-2022

| País | Encuesta | Institución | Año | Identificación de personas migrantes | Desagregación por origen nacional | Identificación de migrantes recientes | Frecuencia de microdatos |
|-------------|--|---|------|--------------------------------------|-----------------------------------|--|--------------------------|
| Argentina | Encuesta Permanente de Hogares | Instituto Nacional de Estadística y Censos | 2020 | Sí | Sí | Sí (país de residencia 5 años antes) | Trimestral |
| | | | 2021 | | | | |
| Bolivia | Encuesta de Hogares | Instituto Nacional de Estadística | 2020 | Sí | No | No | Anual |
| | | | 2021 | | | | |
| Brasil | Pesquisa Nacional por Amostra de Domicílios Contínua | Instituto Brasileiro de Geografia e Estatística | 2020 | No | No | No | Trimestral |
| | | | 2021 | | | | |
| Chile | Encuesta de Caracterización Socioeconómica Nacional | Ministerio de Desarrollo Social y Familia | 2020 | Sí | Sí | Sí (país de residencia 5 años antes) | Anual |
| | | | 2021 | | | | |
| | | | 2022 | Sí | Sí | Sí (país de residencia 5 años antes) | Anual |
| Colombia | Gran Encuesta Integrada de Hogares | Departamento Administrativo Nacional de Estadística | 2020 | Sí | Sí, parcialmente ¹¹ | Sí (país de residencia 1 y 5 años antes) | Mensual |
| | | | 2021 | Sí | Sí, parcialmente ⁷ | Sí (país de residencia 1 y 5 años antes) | |
| | | | 2022 | Sí | Sí | Sí (país de residencia 1 y 5 años antes) | |
| Costa Rica | Encuesta Nacional de Hogares | Instituto Nacional de Estadística y Censos | 2020 | Sí | Sí | Sí (país de residencia 2 años antes) | Anual |
| | | | 2021 | | | | |
| Ecuador | Encuesta Nacional de Empleo, Subempleo y Desempleo | Instituto Nacional de Estadística y Censos | 2020 | No | No | No | Mensual desde setiembre |
| | | | 2021 | Sí | Sí | No | Mensual |
| | | | 2022 | Sí | Sí | No | Mensual |
| El Salvador | Encuesta de Hogares de Propósitos Múltiples | Dirección General de Estadística y Censos | 2020 | No | No | No | N/A |
| | | | 2021 | | | | |
| | | | 2022 | No | No | No | N/A |

¹¹ Solo se presenta la desagregación para algunos orígenes: Estados Unidos, España, Venezuela, Ecuador, Panamá, Perú, Costa Rica, Argentina, Francia e Italia.

| | | | | | | | |
|----------------------|--|--|------|-------------------------------|--------------------------------|--------------------------------------|--------------------------------------|
| Guatemala | Encuesta Nacional de Empleo e Ingresos | Instituto Nacional de Estadística | 2020 | | | | |
| | | | 2021 | No | No | No | Anual |
| | | | 2022 | No ¹² | No | No | Anual |
| Haití | | | | | | | |
| Honduras | Encuesta Permanente de Hogares de Propósitos Múltiples | Instituto Nacional de Estadística | 2020 | | | | |
| | | | 2021 | Sin información ¹³ | Sin información | Sin información | N/A |
| México | Encuesta Nacional de Ocupación y Empleo | Instituto Nacional de Estadística y Geografía | 2020 | | | | |
| | | | 2021 | Sí | Sí, parcialmente ¹⁴ | Sí (país de residencia 1 año antes) | Trimestral ¹⁵ |
| Nicaragua | Encuesta Continua de Hogares | Instituto Nacional de Información de Desarrollo | 2020 | | | | |
| | | | 2021 | Sin información | Sin información | Sin información | N/A |
| Panamá | Encuesta de Propósitos Múltiples | Instituto Nacional de Estadística y Censo | 2020 | | | | |
| | Encuesta de Mercado Laboral Telefónica | | 2021 | Sin información | Sin información | Sin información | N/A |
| Paraguay | Encuesta Permanente de Hogares | Instituto Nacional de Estadística | 2020 | | | | |
| | | | 2021 | Sí | Sí | Sí (país de residencia 5 años antes) | Trimestral |
| Perú | Encuesta Nacional de Hogares | Instituto Nacional de Estadística e Informática | 2020 | | | | |
| | | | 2021 | Sí | Sí | Sí (país de residencia 5 años antes) | Anual |
| Puerto Rico | Puerto Rico Community Survey | United States Census Bureau | 2020 | | | | |
| | | | 2021 | Sí | Sí | Sí (país de residencia 1 año antes) | N/A |
| República Dominicana | Encuesta Nacional de Hogares de Propósitos Múltiples | Ministerio de Economía, Planificación y Desarrollo | 2020 | | | | |
| | | | 2021 | No | No | No | Datos anuales disponibles en REDATAM |
| | | | 2022 | Sí | Sí | No | Datos anuales disponibles en REDATAM |

¹² No se pregunta sobre el lugar de nacimiento de las personas, tal como se realiza en el resto de las encuestas nacionales en las que es posible identificar la condición migratoria. No obstante, en la pregunta “¿Usted se considera o autoidentifica como perteneciente a alguno de los siguientes pueblos?” es posible optar por la categoría de respuesta “extranjero”.

¹³ Si bien no se encuentran disponibles los cuestionarios ni bases de datos de la encuesta para los tres años de interés, siguiendo el relevamiento realizado por Perdomo Rico (2022) y la encuesta de junio de 2023, es posible que sí se pregunta por el lugar y país de nacimiento y el tiempo de residencia en Honduras.

¹⁴ El país de origen solo se presenta desagregado para Estados Unidos, Guatemala y España.

¹⁵ Para el segundo trimestre de 2020 se deben utilizar los datos de la Encuesta Telefónica de Ocupación y Empleo (ETOE).

| | | | | | | | |
|-------------------|--|---|------|-----------------|-----------------|--------------------------------------|-----------|
| Trinidad y Tobago | Continuous Sample Survey of Population | Central Statistical Office - Ministry of Planning and Development | 2020 | | | | |
| | | | 2021 | Sin información | Sin información | Sin información | N/A |
| | | | 2022 | | | | |
| Uruguay | Encuesta Continua de Hogares | Instituto Nacional de Estadística | 2020 | No | No | No | Anual |
| | | | 2021 | Sí | Sí | Sí (país de residencia 5 años antes) | Semestral |
| | | | 2022 | Sí | Sí | Sí (país de residencia 5 años antes) | Mensual |
| Venezuela | Encuesta de Hogares por Muestreo | Instituto Nacional de Estadística | | | | | |

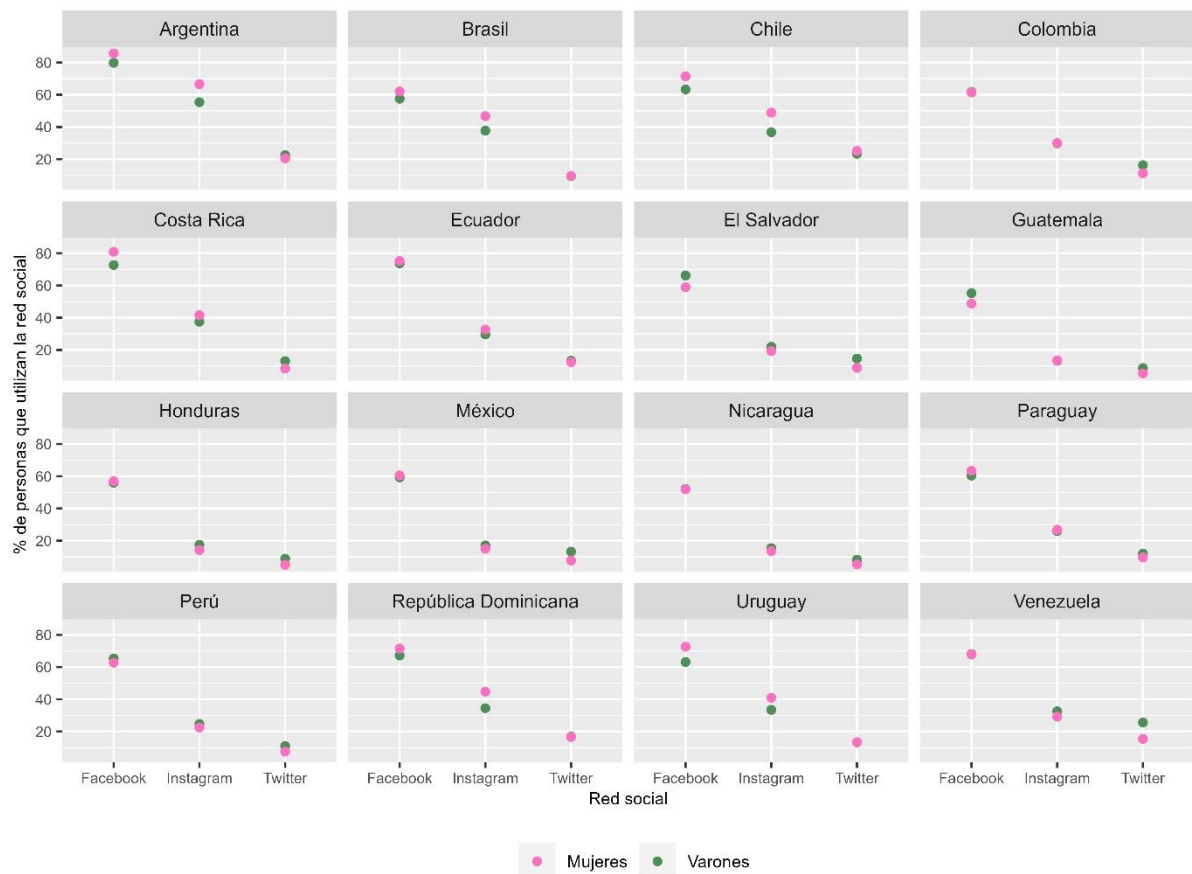
Nota: la etiqueta “sin información” refiere a casos donde se encuentran disponibles informes o cuadros de resultados sobre las encuestas, pero no fue posible acceder a los microdatos, cuestionarios y/o diccionario de variables; N/A=no disponible.

■ Significa que no hay datos disponibles de la encuesta para ese año.

■ Para el caso de Haití, no se encontró una encuesta de referencia salvo la Encuesta sobre Condiciones de Vida de los Hogares tras el terremoto de 2012, mientras que para Venezuela no hay información relativa a la implementación de la Encuesta de Hogares por Muestreo en los años de interés.

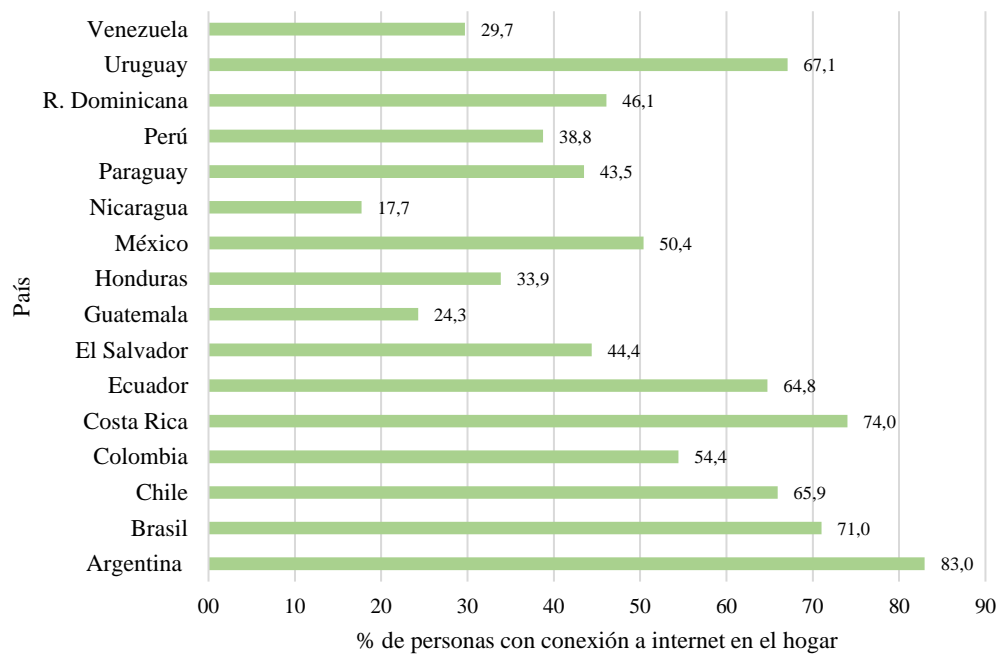
Fuente: elaboración propia a partir de metadatos y diccionario de variables de las encuestas de hogares.

Gráfico 7. Porcentaje de personas usuarias de redes sociales según tipo de red social por sexo. Países y redes sociales seleccionadas, 2020



Fuente: elaboración propia a partir de Latinobarómetro 2020.

Gráfico 8. Porcentaje de personas con conexión a internet en el hogar según país de residencia. Países seleccionados, 2020



Fuente: elaboración propia a partir de Latinobarómetro 2020.