

Ingreso Estructural Por Área Geográfica: una aplicación para Argentina

Nicolás Abbate

CEDLAS-FCE-UNLP

Leonardo Gasparini

CEDLAS-FCE-UNLP y CONICET

Pablo Alfredo Gluzmann

CEDLAS-FCE-UNLP y CONICET

Gabriel Montes Rojas

IIEP-UBA y CONICET

Iván Sznajder

FCE-UNLP

Tobías Yatche

FCE-UNLP

Agosto 2023

Resumen

El objetivo de este trabajo es obtener estimaciones del ingreso estructural para Argentina con un alto nivel de desagregación geográfica, específicamente a nivel de los más de 50.000 radios censales. Para esto estimamos una serie de modelos para el ingreso per cápita familiar en función de características observables para todas las ondas disponibles de la Encuesta Permanente de Hogares en su versión continua (2003-2022) y generamos predicciones del ingreso utilizando las características observables de los hogares en los censos 2001 y 2010. Argentina ha experimentado fuertes vaivenes económicos durante los últimos 20 años, lo que permite obtener predicciones de ingreso bajo distintos estados de la naturaleza. Al incluir todas las estimaciones, podemos predecir el ingreso estructural, entendido como un concepto de mediano plazo donde son los factores estructurales y no los coyunturales, los que tienen mayor influencia. La construcción de esta clase de mapas tiene una importante gama de aplicaciones, y su precisión, desagregación y temporalidad puede ser mejorada utilizando técnicas de inteligencia artificial sobre las imágenes satelitales de las zonas representadas.

JEL: J31, D72

1. Introducción

La tarea de recopilar información sociodemográfica se realiza principalmente a través de encuestas a hogares. Estas herramientas son ampliamente utilizadas en todo el mundo para medir ingresos, pobreza y otros indicadores de bienestar. Realizar cada entrevista es costoso, los cuestionarios necesarios para relevar ingresos o consumo son extensos y requieren una adecuada formación de los encuestadores, por lo que las muestras son relativamente pequeñas con relación a la población de referencia. Esto hace que las mediciones, si bien fiables a nivel de grandes agregados (por ejemplo a nivel nacional, regional, o de aglomerados urbanos), carezcan de representatividad cuando se busca un mayor nivel de desagregación geográfica.

En Argentina, estos problemas se acentúan: la Encuesta Permanente de Hogares (EPH), de periodicidad trimestral, es la principal fuente de información utilizada para medir ingresos, pobreza, desigualdad y otros indicadores socioeconómicos relevantes. Debido a los altos costos, su cobertura abarca solamente a los principales aglomerados urbanos. La EPH Ampliada (o Encuesta Anual de Hogares Urbanos) extiende la cobertura a la toda la población urbana. Pero, en ambos casos, la desagregación espacial máxima (o nivel de representatividad) es a nivel de aglomerado (31 aglomerados urbanos para la EPH y 22 más considerando la versión ampliada). Adicionalmente, la Encuesta Nacional de Gasto de los Hogares (ENGHo) tiene cobertura nacional, contiene información de ingresos y gastos de los hogares, pero su periodicidad es menor (cada 5 años aproximadamente) y, debido a su muestra limitada, sólo es representativa a nivel de provincias.

Por otro lado, los censos, que se realizan con una frecuencia aproximada de diez años, permiten obtener información de prácticamente todas las personas del territorio. Esto implica una capacidad de desagregación granular: la unidad máxima de desagregación son los radios censales, que agrupan en promedio a 300 viviendas, por lo que en áreas de alta densidad poblacional la información se presenta a nivel de manzana. Sin embargo, en Argentina los censos no recolectan información sobre ingresos ni gastos, por lo que no es posible estimar indicadores de pobreza monetaria de forma directa. Con esta información solo es posible calcular indicadores de pobreza multidimensional que no consideren variables monetarias.

Dados los límites que tienen los datos de las encuestas, la literatura ha ensayado una serie de soluciones que buscan resolver el problema de la falta estimaciones directas espacialmente desagregadas. Estas técnicas suelen denominarse en la literatura como “estimaciones de área pequeña” (*small area estimation*, SAE*). Las técnicas de SAE proveen un marco analítico para mejorar el nivel de granularidad de la información, sin requerir la recolección de una mayor cantidad de datos. En general, estas estimaciones requieren incorporar fuentes de información auxiliares para mejorar su precisión.

*Burke et al. (2021) identifican una serie de desafíos relevantes que se presentan al implementar y utilizar las encuestas de hogares. En primer lugar, las encuestas a nivel nacional son costosas y requieren una gran cantidad de tiempo para su realización. En segundo lugar, las encuestas, en general, presentan, como máximo, desagregación a nivel de aglomerados urbanos, limitando muchas aplicaciones que requieren una mayor desagregación. Por último, los estimaciones generadas con datos de campo pueden ser poco precisas, presentar sesgos, o ser poco representativas de la población.

Uno de los métodos más utilizados de *small area estimation* (SAE) es el método desarrollado por [Elbers et al. \(2003\)](#). Este método requiere de la disponibilidad de los microdatos del censo y de una encuesta, preferentemente contemporáneos. Con el uso de datos de encuestas se construye un modelo para la relación entre una medida del bienestar del hogar, como el gasto per cápita del hogar, y un conjunto de características del hogar, sus miembros y su ubicación. Haciendo uso de los parámetros estimados del modelo, se utilizan las características en el censo de la población del país para simular el gasto per cápita del hogar (o los ingresos) para cada hogar individual. A partir de los valores simulados del censo sobre el bienestar podemos calcular cualquier indicador de bienestar. Esta metodología ha sido utilizada extensivamente, particularmente por organismos nacionales de estadísticas y organizaciones internacionales, para la estimación de mapas de pobreza. Entre muchos ejemplos pueden citarse los mapas de pobreza de Brasil ([Ebers et al., 2008](#)), Vietnam ([Cuong et al., 2010](#)) y Sri Lanka ([Department of Census and Statistics and World Bank, 2015](#)). Desde la publicación de esta metodología ha habido avances considerables en la literatura, como el estimador *Empirical Best* (EB), que además incorpora al modelo información sobre el estimador en la encuesta ([Molina and Rao, 2010](#)).

Adicionalmente al problema de la desagregación espacial, definir la pobreza de manera precisa ha resultado ser un desafío. La métrica más comúnmente empleada se basa en los ingresos, estableciendo que una persona es considerada pobre si sus ingresos familiares caen por debajo de un umbral predefinido. No obstante, esta aproximación puede llevar a identificar erróneamente a individuos con ingresos temporariamente bajos como pobres, distorsionando la imagen real de la pobreza y generando correlaciones espurias con el desempleo.

Sumado a lo anterior, la medición del ingreso como medida del bienestar depende del momento en el cual se releva la información. El ingreso que usualmente relevan las encuestas de hogares es el ingreso corriente; típicamente el del mes previo a la encuesta. Este es el caso para la gran mayoría de los países de América Latina, incluido Argentina, el país elegido para la ilustración de nuestra metodología ([Beccaria and Gluzmann, 2013](#)). Un problema importante, generado por esta limitación, proviene de las fluctuaciones del ingreso a lo largo de un año (o de períodos de tiempo más extensos). En particular, el ingreso en un determinado mes puede ser bajo por razones circunstanciales, como shocks negativos de corto plazo y fluctuaciones estacionales.

Estos inconvenientes proveen el argumento principal a favor de los indicadores basados en consumo, dado que esta dimensión es menos dependiente de circunstancias coyunturales, pero su relevamiento (mediante encuestas) es aún más costoso debido a la cantidad de información a recabar. En Argentina casi no existen fuentes disponibles periódicas que permitan estimaciones de consumo (a excepción de las mediciones parciales basadas en Gasto que pueden obtenerse de las ENGHOS).

Los microdatos de tipo panel permiten medir el ingreso de una persona u hogar a través del tiempo, lo que permitiría evaluar si los ingresos son altos o bajos en el largo plazo. Sin embargo, existen inconvenientes al aplicar esta alternativa: las encuestas de panel rotativo siguen hogares o familias por distintos períodos, pero no mayores a 1 o 2 años, por lo que rango temporal es limitado para considerarlo estructural o “permanente”. A su vez, el grado de desgaste (*attrition*) aumenta con el

tiempo, por lo que un grupo no aleatorio de individuos puede perderse a medida que transcurre el tiempo de relevamiento. En los relevamientos que buscan medir el ingreso en períodos más largos (como Panel CASEN, en Chile) el problema de desgaste se agrava dada la dificultad de ubicar al mismo individuo luego de un plazo considerable. Pero aún sin problemas de desgaste, debemos tener en cuenta que es necesario esperar muchos años para poder obtener una medición. En Argentina, la única encuesta con estructura de panel es la EPH, pero solamente permite seguir parcialmente a los individuos durante 1 año y medio, y se volvió imposible desde 2016 dado que el instituto decidió aleatorizar los identificadores de hogares para las bases usuarias.

El camino elegido en el presente trabajo para identificar a los ingresos de las personas, particularmente de los más pobres, se encuentra a mitad de camino entre un enfoque basado en las características observables de las personas y los hogares (como puede ser el NBI) y el ingreso o consumo corriente. El objetivo de este trabajo es aplicar los desarrollos en términos de SAE para el caso argentino, pero incorporando la dimensión estructural de las estimaciones espaciales del ingreso. Para ello, siguiendo la estrategia propuesta [Gasparini et al. \(2022\)](#) se utilizan todas las ondas disponibles de la Encuesta Permanente de Hogares en su versión continua (en lugar de utilizar únicamente los datos contemporáneos). Con esta información, la estimación del ingreso espacial ya no se corresponderá con una coyuntura económica determinada —por ejemplo, el ingreso medio de los hogares para el 2do semestre de 2010—, sino que tendrá un componente estructural —el ingreso “esperado” de los hogares. En otras palabras, la metodología propuesta permite resolver la condición coyuntural de las estimaciones por de SAE (como ELL o EB), permitiendo considerar el ingreso que los hogares hubieran tenido en situaciones económicas más —o menos— favorables.

El resto del trabajo se organiza de la siguiente forma. La sección 2 presenta la metodología utilizada para la estimación del ingreso estructural. La sección 3 presenta los principales resultados y compara con medidas alternativas de ingresos y bienestar social. Se desarrolla una aplicación y extensión de este método utilizando imágenes satelitales en la sección 4. El trabajo concluye en la sección 5.

2. Metodología

El presente trabajo se centra en la construcción de un índice de ingreso estructural por área geográfica, es decir, un indicador del ingreso espacial para todo el territorio argentino. La construcción de la estimación se realiza estimando el ingreso per cápita familiar para cada individuo del censo desde el 2003 hasta el 2019 y agregando estas estimaciones del ingreso por área geográfica. Se utilizan las Base de Datos Socioeconómicos para América Latina y el Caribe (SEDLAC, por sus siglas en inglés) un proyecto realizado conjuntamente por el Centro de Estudios Distributivos, Laborales y Sociales (CEDLAS) de la Universidad Nacional de La Plata en Argentina y el grupo de Pobreza y Género de América Latina del Banco Mundial (LCSPP). Se incluye la homogeneización del método de imputación de ingresos propuestos en [Tornarolli \(2018\)](#). Los ingresos per cápita familiares se expresan a precios 2011 para la consistencia intertemporal.

Para estimar el ingreso per cápita familiar para cada individuo en cada semestre se utiliza un modelo de regresión condicional basado en la ecuación de Mincer, la cual relaciona los ingresos con el capital humano de las personas y las características de los hogares e individuos. De esta manera es posible evaluar cómo las variables observadas explican el ingreso per cápita familiar para los jefes de hogares, el indicador más importante de los ingresos de los hogares y las personas, en cada onda de la EPH.

Una vez construido el IPCF estimado para cada individuo en cada semestre de la EPH, se construye el ingreso estructural de cada individuo del censo como el promedio de las predicciones del ingreso para el período 2003-2022. Por último, se estima el ingreso estructural por radio censal como la media del ingreso estructural individual para cada radio censal para los últimos dos censos disponibles.

La estimación del ingreso se realizó mediante una variación del método desarrollado por [Elbers et al. \(2003\)](#). En el trabajo original, los autores plantean una estimación de área pequeña que consiste en: (i) seleccionar un indicador de interés que se encuentre disponible en una encuesta pero no en el censo, (ii) identificar covariables que se relacionen con el indicador y que se encuentren tanto en el censo como en la encuesta, (iii) estimar un modelo lineal de mínimos cuadrados generalizados (GLS) sobre la encuesta, que relacione el indicador de interés con las covariables seleccionadas, y (iv) generar predicciones del indicador en el censo, utilizando las covariables censales disponibles y las estimaciones de los parámetros del modelo del punto previo. Estas predicciones tienen la ventaja de estar geográficamente localizadas, ya que típicamente los censos permiten una desagregación espacial mucho mayor que las encuestas de hogares. En general, luego de esta predicción se agregan los indicadores al mínimo nivel espacial disponible, lo que permite construir mapas con las estimaciones correspondientes. Algunas contribuciones importantes a la precisión de la estimación son las propuestas por [Molina and Rao \(2010\)](#) y CENSUSEB, revisar.

Para realizar la estimación espacial del ingreso, entonces, se estimó un modelo lineal sobre los microdatos de cada uno de los semestres de la Encuesta Permanente de Hogares entre 2003 y 2022, estimando la relación entre el ingreso per cápita familiar de cada hogar i del aglomerado j (y_{ij}^t) y un vector de variables observables (X_{ij}^t). Las variables incorporadas son la edad, el género, binarias para rangos etarios, y características relacionadas con la vivienda, como disponibilidad de agua potable, acceso a baño y condición de propiedad. También se ha incluido información sobre el nivel educativo máximo alcanzado y la cantidad de miembros en el hogar. Además, se han considerado variables específicas del jefe o jefa del hogar, como edad, nivel educativo y género. En el modelo se incluyen efectos fijos por aglomerado δ_j^t . Por lo tanto, para el semestre t de la Encuesta Permanente de Hogares y el censo s , se estimó un modelo correspondiente es de la forma:

$$y_{ij}^t = \beta^t X_{ij}^t + \delta_j^t + \varepsilon_{ij}^t \quad (1)$$

Una vez estimado el modelo, se guardan las estimaciones del vector de parámetros de cada una de las ondas β^t y δ_j^t . Utilizando estas estimaciones, se generan predicciones del ingreso para cada uno de

los hogares del censo, utilizando las covariables de los últimos dos censos disponibles X_{icj} . En otras palabras, se utilizan la relación estimada en la EPH entre las variables observables de los hogares y el ingreso en cada EPH (t), y se utiliza esa relación para predecir el ingreso utilizando las variables observables del censo. La única diferencia es que, en el caso del censo, el hogar i además tiene una identificación con un radio censal c utilizando el censo s , que será posteriormente utilizado para la construcción del índice de ingreso. Específicamente, se computa la siguiente relación:

$$\hat{y}_{icj}^{ts} = \beta^t X_{icj}^s + \delta_j^t \quad (2)$$

Esta relación, estimada para cada semestre de la EPH, puede interpretarse como el ingreso que el hogar i del radio censal c del aglomerado j encuestado en cada uno de los censos s tendría bajo la situación económica correspondiente al trimestre t . En este sentido, las variaciones de los rendimientos de cada una de las características observables estimadas según la ecuación 1 a lo largo del tiempo se entienden como cambios en la situación económica. Para los hogares rurales se imputó el efecto fijo del aglomerado más cercano, ya que no existen datos de ingresos de personas que viven en la ruralidad en la EPH.

El *ingreso estructural* del individuo i del Censo s (\bar{y}_{icj}^s) se computa promediando las predicciones del ingreso correspondientes a cada una de las ondas t de la EPH y Censo S . Esto es:

$$\bar{y}_{icj}^s = \sum_{t \in T} \frac{\hat{y}_{icj}^{st}}{n_T} \quad (3)$$

donde T representa el conjunto de ondas de la EPH y n_T el tamaño de ese conjunto. Este ingreso estructural puede interpretarse como el ingreso que el hogar i del radio censal c del aglomerado j tuvo en una situación económica promedio utilizando datos del censo s . La ventaja de esta estimación por sobre métodos tradicionales de *small area estimation* es que permite considerar un lapso de tiempo mayor a la coyuntura del semestre de la encuesta seleccionada, permitiendo identificar los hogares que *estructuralmente* tienen ingresos altos o bajos.

Finalmente, para construir el indicador de *ingreso estructural* por radio censal se calcula la media del ingreso estructural individual para cada uno de los radios censales. Por lo tanto, definiendo n_c como el número de hogares contenidos en el radio censal c , el *ingreso estructural* del radio censal c (\bar{Y}_{cj}^s) queda definido como:

$$\bar{Y}_{cj}^s = \sum_{i \in c} \frac{\bar{y}_{icj}^s}{n_j} \quad (4)$$

Con esta estimación, resulta posible generar mapas del ingreso estimado por radio censal para, de ese modo, tener una estimación del ingreso geográfico para todos los radios censales disponibles en el Censo 2010.

A diferencia del método planteado por (Elbers et al., 2003), en lugar de realizar una estimación por mínimos cuadrados generalizados, al estimar el modelo 1 se utiliza mínimos cuadrados ordinarios. Esta diferencia se debe a que los autores cuentan con una encuesta que identifica a los hogares con la mínima unidad de agregación del censo. En cambio, utilizando la EPH, no se dispone, en los datos, de información sobre a qué radio censal pertenece cada hogar, por lo que no es posible estimar la matriz de varianzas y covarianzas necesarias para una estimación de GLS. Utilizar el método de mínimos cuadrados ordinarios resulta en una estimación sesgada del ingreso de los hogares, ya que resulta esperable que los residuos ε se correlacionen dentro de un mismo radio censal. A pesar de esto, se espera que el resultado de las estimaciones presenten un sesgo mitigado. Esto se debe a que, si bien las estimaciones de cada hogar estarán sesgadas, se espera que la agregación del ingreso por radio censal mitigue el sesgo en cierta medida, ya que los errores no tienen una correlación perfecta y en buena medida se contrarrestarían.

3. Resultados

3.1. Precisión de las estimaciones sobre la EPH

En esta sección nos concentramos en evaluar la performance de las predicciones del modelo de ingreso estructural (IE) en comparación con el ingreso per cápita familiar (IPCF) observado en la encuesta. Antes de analizar los resultados resulta útil establecer ciertas hipótesis respecto de dicho desempeño. Podemos separar las discrepancias entre IE e IPCF en “deseables” e “indeseables”.

Las discrepancias deseables están relacionadas con el carácter estructural del IE y el carácter corriente del IPCF. En términos de variabilidad, deberíamos esperar una menor dispersión (y desigualdad) del IE, puesto que, en la media que el IE aproxime mejor el concepto “estructural”, éste indicador debería verse menos afectado por variaciones coyunturales que el IPCF. En términos de sesgo, podrían presentarse diferencias si el ingreso corriente presentara mayores o menores niveles en algún grupo no aleatorio de hogares, pero en este caso el sesgo no tiene a priori un signo definido.

Las discrepancias “indeseables” se relacionan con la falta de información (i.e inobservables de carácter estructural potencialmente relevantes), así como la incorrecta especificación del modelo econométrico. En ambos casos las discrepancias pueden provocar inconsistencias tanto en términos de sesgo como de varianza. El carácter lineal del método de estimación (MCO) elegido puede provocar sesgos por error de especificación en la medida que se presenten no linealidades entre el ingreso y las variables explicativas. Adicionalmente, las covariables elegidas ad-hoc pueden no ser suficientes para predecir

Cuadro 3.1: Estadísticas descriptivas del Ingreso estructural y el ingreso per cápita familiar.

	Ingreso Estructural (predicho)	Ingreso per cápita familiar (observado)
Media	1.447,0	1.847,5
Mediana	1.190,1	1.374,7
Media por quintiles	1	520,7
	2	859,3
	3	1.199,7
	4	1.705,0
	5	2.950,2
Coefficiente de gini	0,336	0,436

Fuentes: IPCF calculado utilizando la EPH del 2do semestre de 2010, Ingreso Estructural calculado utilizando características de la EPH 2do semestre de 2010 y modelos estimados para todas las ondas de EPH 2003-2022.

correctamente la relación. La razón por la cual se aplica esta estrategia es simplemente operativa, ya que en futuras versiones se buscará evaluar métodos más sofisticados de estimación (métodos semi paramétricos), así como incorporar de distintos métodos de selección de variables relevantes, como por ejemplo la inclusión más variables disponibles tanto en la EPH como en el censo, la incorporación de interacciones, la selección de variables mediante *lasso* (Tibshirani, 1996), Global Search Regression (Gluzmann and Panigo, 2015; Panigo et al., 2020, 2023), entre otros. El modelo elegido debería subestimar la variabilidad, dado que parte del error incluye variabilidad de carácter estructural, y el método carácter lineal probablemente genere sesgos en las colas de la distribución. En términos de desigualdad, el IE presenta un coeficiente de Gini sustancialmente menor, aunque es difícil a priori establecer en qué medida esto se debe a discrepancias deseables o indeseables.

El Cuadro 3.1, muestra un menor valor del IE para medidas centrales (media y mediana). Una posible hipótesis que explique este resultado se relaciona con que el año 2010 fue un año relativamente “bueno” para Argentina. Al observar las diferencias por niveles de ingreso, se evidencia un mayor nivel respecto del IPCF para el quintil más bajo, y menores niveles para el resto de los quintiles. Sin embargo, esta diferencia es creciente a medida que se incrementa el nivel de ingresos, especialmente en el 5to quintil. Si bien parte de estas discrepancias pueden ser de carácter “deseable”, es probable que éstas sean principalmente producto de diferencias indeseables, especialmente en el tramo de mayores ingresos.

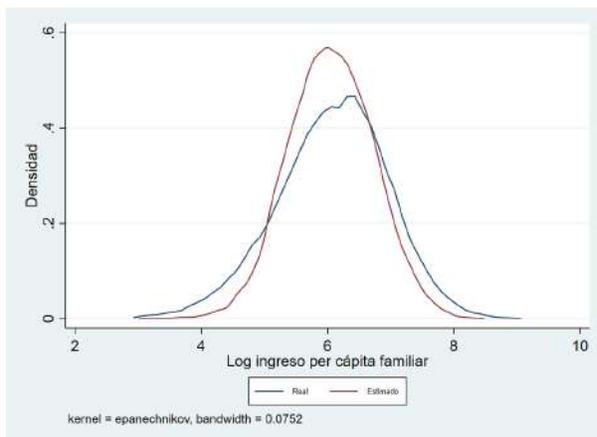
La Figura 1 permite visualizar las diferencias antes mencionadas a lo largo de toda la distribución. El panel (a) muestra estimaciones de kernel del logaritmo del IE y el IPCF. Se observa que existen diferencias a favor del IPCF en el centro de la distribución, pero las curvas difieren en las colas, con mayor concentración del IPCF en estos tramos. El panel (b) compara, a nivel de individuo, el IPCF con su IE. Puede observarse que existe una mayor dispersión de los puntos en los niveles más bajos de la distribución del ingreso y una mayor masa a la derecha de la curva de 45 grados en el tramo de mayores ingresos.

Cuadro 3.2: Ingreso estructural e IPCF, por regiones

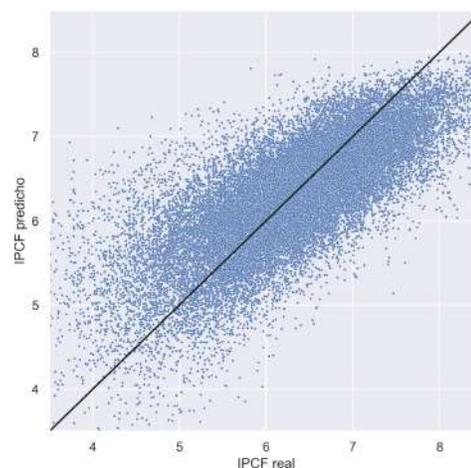
	Ingreso Estructural (predicho)	Ingreso per cápita familiar (observado)
Gran Buenos Aires	1.586,8	2.063,9
Pampeana	1.424,5	1.788,0
Cuyo	1.175,3	1.349,6
Noroeste Argentino	1.026,9	1.271,1
Noreste Argentino	977,0	1.166,3
Patagonia	1.918,8	2.606,5
Total	1.447,0	1.847,5

Fuentes: IPCF calculado utilizando la EPH del 2do semestre de 2010, Ingreso Estructural calculado utilizando características de la EPH 2do semestre de 2010 y modelos estimados para todas las ondas de EPH 2003-2022.

Figura 1: Distribución IPCF e IE



(a) Comparación del IPCF predicho y el real



(b) Distribución real y estimada del IPCF

El Cuadro 3.2 muestra las diferencias a nivel de regiones del país. En todas las regiones se observan mayores ingresos en términos del IPCF, siendo las regiones de GBA y La Pampa (regiones con ingresos mayores al promedio nacional) las que presentan mayores diferencias.

Finalmente, el Cuadro 3.3 muestra distintas variables asociadas al carácter estructural de los ingresos. En la medida en que el IE represente mejor niveles de ingreso de mediano/largo plazo, deberíamos esperar que los individuos de menores ingresos presenten peores desempeños en términos de vivienda e infraestructura, años educación y necesidades básicas insatisfechas. En el cuadro siguiente se observan efectivamente estos patrones para el quintil de menores ingresos. En la última fila se muestra el porcentaje de desocupados respecto de la población económicamente activa, una variable predominantemente de carácter coyuntural. Se observa que la diferencia se revierte respecto de las variables anteriores, lo que es “deseable” para un indicador que busca representar niveles de ingreso de mediano/largo plazo.

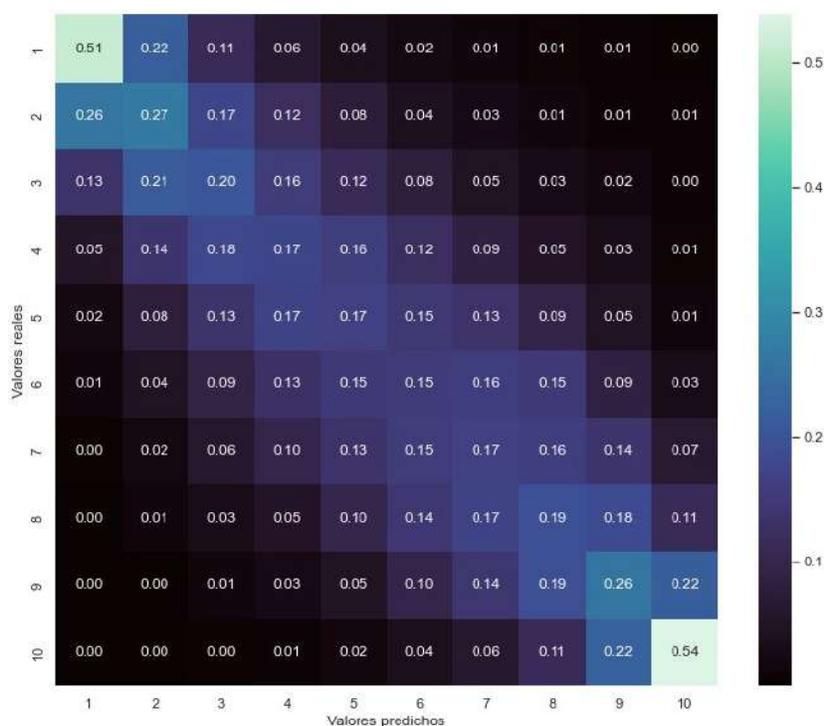
Cuadro 3.3: Ingreso estructural (IE) e Ingreso per capital familiar (IPC) en relación a características observables de los hogares

	Quintiles de ingreso estructural y observado									
	1		2		3		4		5	
	IE	IPC	IE	IPC	IE	IPC	IE	IPC	IE	IPC
Vivienda en lugar precario (%)	9.0	6.5	3.3	3.4	2.2	2.2	0.9	1.1	0.6	0.6
Vivienda con mat. precarios (%)	7.3	5.3	2.4	2.6	1.3	1.3	0.6	0.9	0.3	0.2
Vivienda con acceso a agua (%)	95.9	97.3	99.3	98.9	99.2	99.0	99.6	99.7	100	99.9
Vivienda con baño (%)	54.5	71.0	87.7	85.2	93.1	91.9	96.7	96.5	99.8	99.0
Vivienda con acceso a cloaca (%)	21.7	37.1	40.6	50.0	55.6	56.4	70.3	68.9	87.6	84.6
Hogares con NBI (%)	69.7	51.9	32.1	31.5	18.4	22.1	12.5	11.8	4.5	4.1
Miembros por habitación	1.61	1.24	0.98	0.94	0.75	0.73	0.56	0.58	0.38	0.44
Años de educación promedio	5.67	6.18	7.36	7.34	8.60	8.32	9.75	9.55	12.1	12.1
Desempleo (% personas activas)	11.0	17.6	8.9	10.8	7.9	7.8	6.9	5.5	4.3	2.1

Fuentes: IPCF calculado utilizando la EPH del 2do semestre de 2010, Ingreso Estructural calculado utilizando características de la EPH 2do semestre de 2010 y modelos estimados para todas las ondas de EPH 2003-2022.

La Figura 2 muestra la correlación por deciles entre el IPCF y el IE. Se observa que, en general, el modelo capta mejor los extremos de la distribución, mientras que tiene más problemas para captar los deciles medios de manera correcta. Sin embargo, se observa que cuando el modelo predice “incorrectamente” el decil, lo hace por un valor cercano al real.

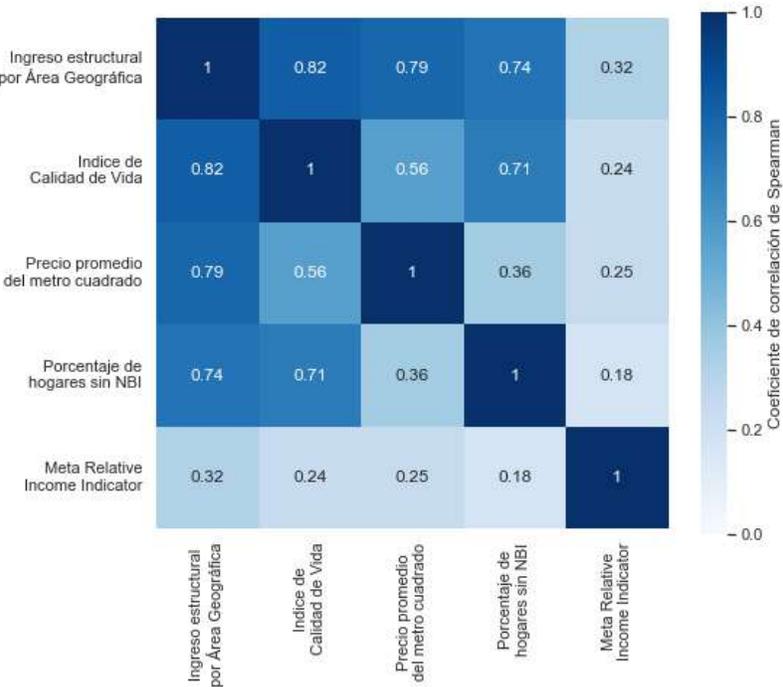
Figura 2: Correlación por deciles entre el IPCF y el IE



Nota: Valores normalizados respecto a los deciles reales

La Figura 3 se presenta un correlograma entre el IE y otros estimadores del bienestar. Se observa que, en términos generales, la correlación es alta, particularmente entre el ingreso estructural y los indicadores de bienestar que, a su vez, no se correlacionan entre sí. Lo anterior permite interpretar al IE como una variable latente —el bienestar— que expresan los restantes indicadores.

Figura 3: Correlación entre IE y otras medidas de bienestar



3.2. Ingreso estructural por Área Geográfica

A partir de la metodología descrita en las secciones anteriores es posible construir mapas del Ingreso estructural por radio censal utilizando los dos últimos censos disponibles.

En la Figura 4 se presenta la distribución relativa del Ingreso Estructural por Área Geográfica para la Argentina utilizando las características observables del censo 2001 y 2010. En ambos casos se muestra el decil del IEPAG correspondiente a cada radio censal. En términos generales, ambos mapas presentan un patrón similar: el sur de la Argentina tiene un ingreso medio-alto, el centro un ingreso medio que decrece hacia el este, y el norte presenta un ingreso bajo que decrece en la región del NOA. En términos temporales, sin embargo, hubo un empobrecimiento relativo en el interior de la Provincia de Buenos Aires, San Luis, Córdoba y Santa Fe entre el 2001 y 2010. Sin embargo, en el resto del país la pobreza relativa se mantiene constante entre ambos períodos. Esto se explica por el crecimiento relativo del ingreso en los centros urbanos.

Al comparar la distribución del ingreso estructural se observa un corrimiento a la derecha para el año 2010. El ingreso estructural por radio censal mejoró entre el año 2001 y 2010. Esto puede observarse en la Figura 5. El corrimiento de la distribución completa a la derecha refleja un cambio, no en el ingreso “coyuntural”, sino una mejoría en las características observables de los hogares: mayor educación promedio, mejores condiciones de vivienda, y una mejoría en todas las características utilizadas en la predicción.

Figura 4: Ingreso estructural por Área Geográfica, evolución de la distribución relativa

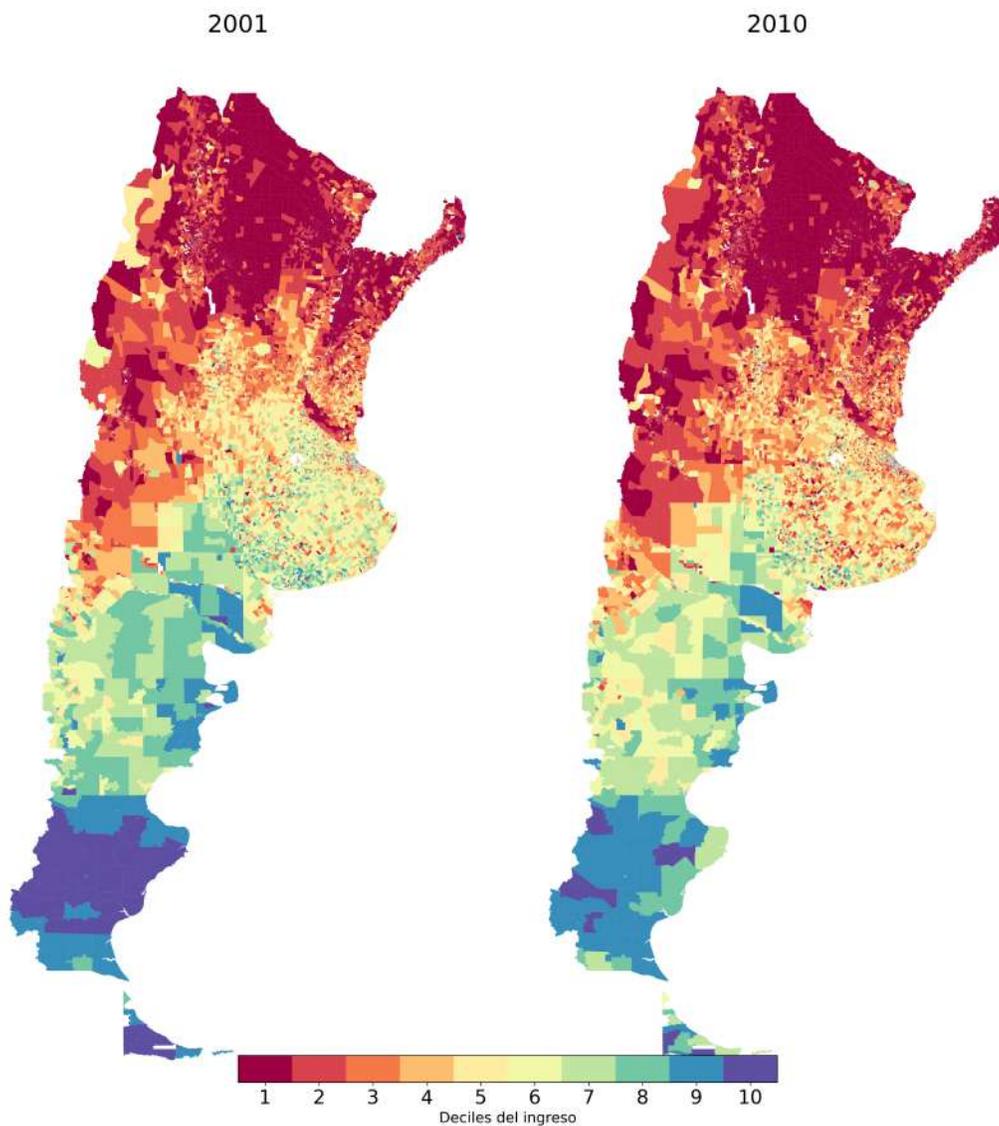
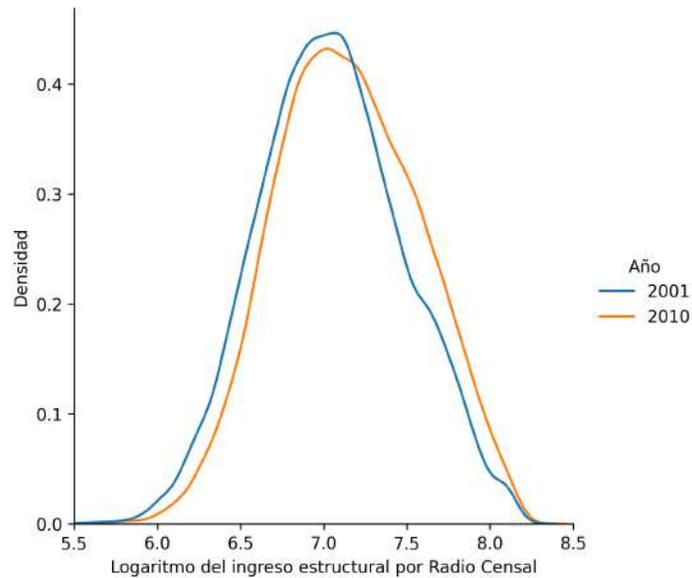


Figura 5: Distribución Logaritmo Ingreso estructural 2001 y 2010

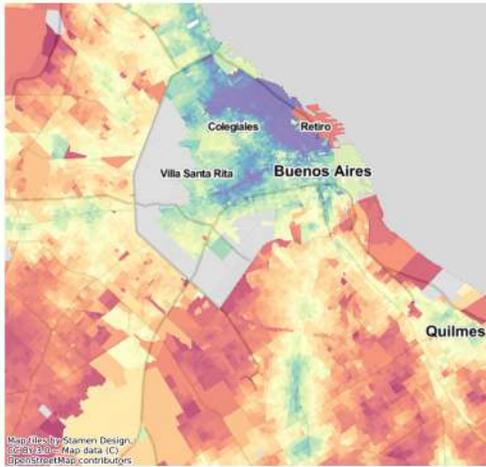


Además de estudiar lo que sucede con la distribución relativa del ingreso estructural, resulta posible, a partir de la metodología aplicada en el presente trabajo, analizar la distribución absoluta del ingreso estimado.

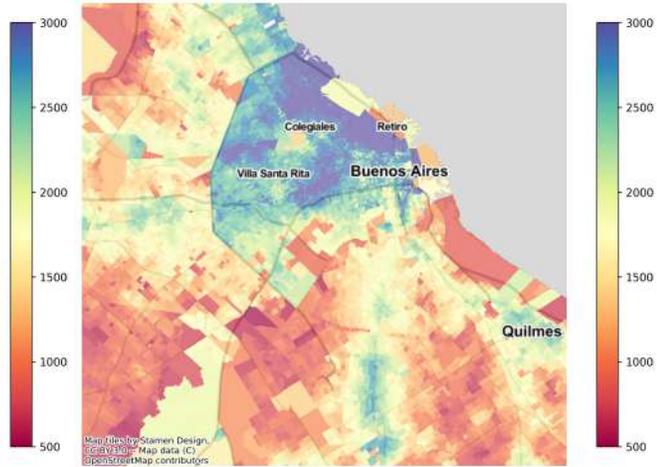
En la Figura 8 se observa que, en las principales zonas urbanas del país, el ingreso crece en términos absolutos entre los años 2001 y 2010. Al igual que nivel nacional, este fenómeno se da por una mejora en las características observables de los individuos. Sin embargo, se observa que en La Plata no hay cambios relevantes entre ambos censos.

La Figura 7 presenta la evolución relativa del ingreso en las capitales seleccionadas. Para ello, clasifica en quintiles la diferencia del ingreso entre 2001 y 2010 para estas ciudades; es decir, pertenecer al quintil 2 significa que el hogar se encuentra entre el 20% y el 40% de los hogares que menos creció en la ciudad. Tal como se mencionó y puede observarse en las etiquetas de estos quintiles, la tendencia de las ciudades entre 2001 y 2010 presenta un crecimiento absoluto: solo el primer quintil presenta un empobrecimiento absoluto, mientras que los cuatro restantes tienen crecimiento a diferente tasa. Por otra parte, se observa que, en términos relativos, en las zonas centrales de las principales zonas urbanas del país el crecimiento del ingreso estructural fue menor que en las zonas aledañas.

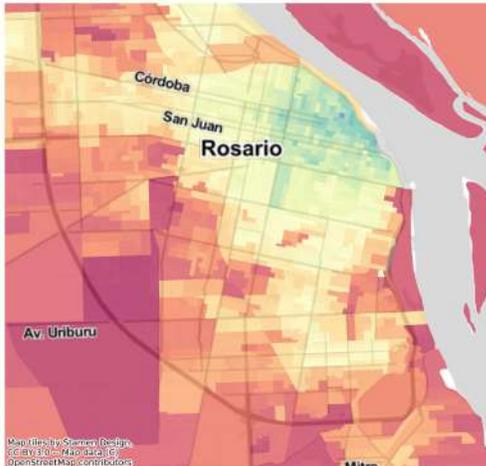
Figura 7: Ingreso estructural por Área Geográfica, evolución de la distribución absoluta



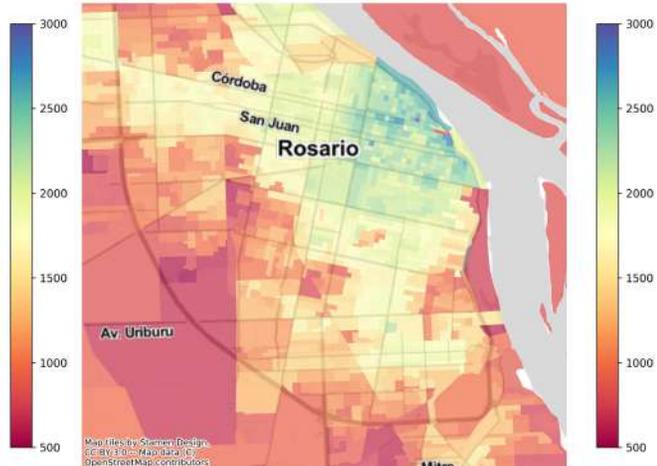
(a) AMBA, 2001



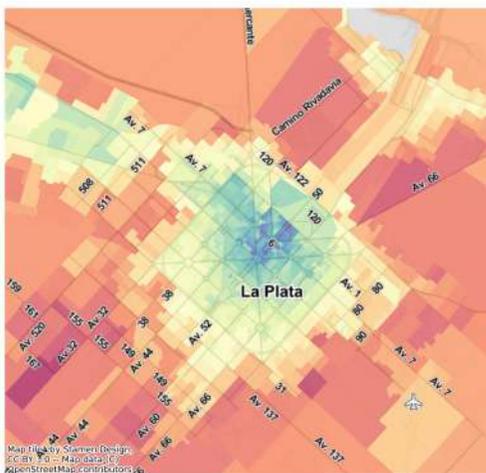
(b) AMBA, 2010



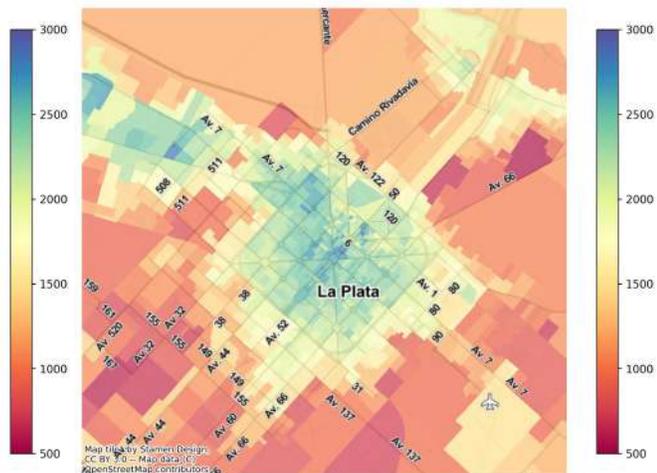
(c) Rosario, 2001



(d) Rosario, 2010

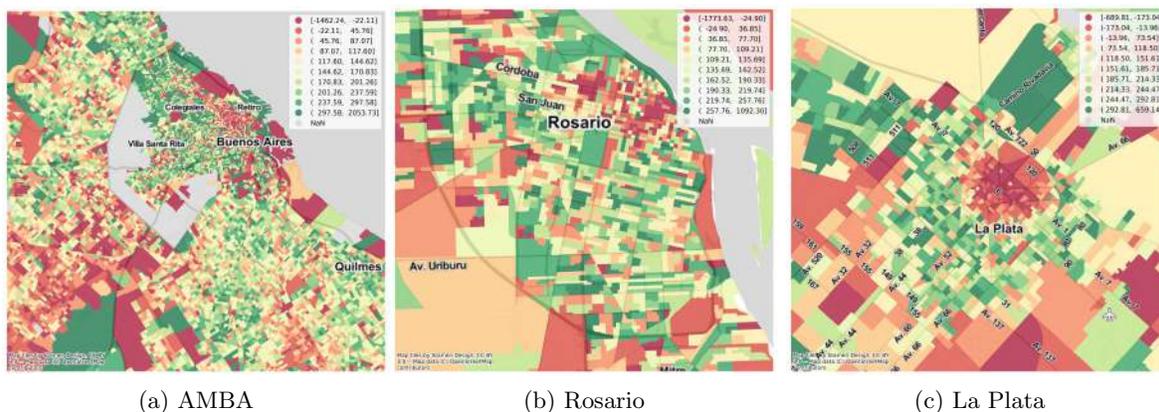


(e) La Plata, 2001



(f) La Plata, 2010

Figura 6: Deciles de la diferencia del Ingreso estructural por Área Geográfica entre 2001 y 2010



4. Imágenes Satelitales para mejorar las estimaciones

A pesar de que las estimaciones de áreas pequeñas ofrecen la ventaja de mejorar significativamente la granularidad y suministrar información minuciosa sobre regiones específicas, persiste una problemática considerable cuando se basan exclusivamente en datos provenientes de encuestas. La cuestión radica en que estas estimaciones solo pueden crearse con la misma periodicidad que las fuentes utilizadas. En situaciones donde se emplean datos censales, como en el caso de la metodología aplicada a lo largo de este trabajo, los indicadores espaciales solo pueden generarse cada diez años. Como resultado, las estimaciones no logran captar adecuadamente los cambios recientes en la población o en las condiciones locales, lo cual puede limitar su aplicabilidad en la toma de decisiones. Por otra parte, por su mecanismo de generación, las estimaciones de ingreso se limitan únicamente al ámbito nacional, impidiendo comparaciones entre países.

Con el propósito de brindar estimaciones espaciales más actualizadas y detalladas del bienestar, recientemente han surgido enfoques de Estimación de Áreas Pequeñas (SAE) que hacen uso de datos no convencionales. Estos enfoques engloban el empleo de datos de "luces nocturnas" captadas por satélites (Henderson et al., 2012; Ciaschi, 2021), información de teléfonos móviles (Blumenstock et al., 2015; Steele et al., 2017), datos de redes sociales (Weber et al., 2018; Sheehan et al., 2019), imágenes satelitales de alta resolución (Jean et al., 2016; Yeh et al., 2020; Engstrom et al., 2022), e incluso combinaciones de estos recursos. Gran parte de estos estudios recurren a técnicas basadas en aprendizaje automático (*machine learning*) para procesar datos no estructurados, como imágenes en el caso de información satelital, o vastos volúmenes de texto en el contexto de las redes sociales. Por lo general, estos estudios entrenan un modelo de aprendizaje automático para que emplee esta información no estructurada como predictor y genere predicciones de alguna medida de bienestar basada en encuestas (frecuentemente desarrollada mediante el método ELL). Una vez entrenado el modelo, se utiliza para producir estimaciones de áreas pequeñas en todo el ámbito de interés, lo que posibilita la creación de mapas del

indicador de interés.

La generación de estimaciones de áreas pequeñas mediante técnicas de aprendizaje automático es un campo que ha tomado relevancia recientemente, impulsado por los avances en las capacidades de análisis de las redes neuronales artificiales y el notable incremento en la capacidad de cálculo de las tarjetas gráficas. Usualmente, estas estimaciones son muy precisas y resistentes a posibles fallos en los datos de entrada (Burke et al., 2021). Además, la información auxiliar tiende a estar disponible con alta frecuencia. Por ejemplo, existen diversos satélites que proveen imágenes de alta resolución y sin nubes con intervalos temporales inferiores a un año. Esto subraya el potencial de aprovechar fuentes alternativas que guarden correlación con los indicadores de interés para llenar los vacíos temporales entre censos en los mapas de información.

A pesar de que ampliar los métodos convencionales de recopilación de datos sería costoso, en la actualidad se recogen con mayor frecuencia otros tipos de datos que podrían ser pertinentes para medir el bienestar y otros indicadores sociodemográficos. Los avances tecnológicos recientes han incrementado la disponibilidad global de imágenes satelitales de alta resolución, las cuales contienen información relevante sobre características que podrían estar correlacionadas con varios de estos indicadores. Por ejemplo, densidades de población altas o bajas, así como niveles de ingresos altos, medios o bajos, pueden ser identificados visualmente en estas imágenes (ver Figura 8).

Recientemente, las imágenes satelitales de alta resolución han surgido como una fuente práctica de información sobre el bienestar, en gran parte gracias a los avances en los algoritmos de visión por computadora. Los progresos en el aprendizaje profundo, como las Redes Neuronales Convolucionales (CNN), permiten clasificar imágenes basándose en patrones comunes identificados internamente para cada categoría (clasificación o regresión de imágenes). Además, estos algoritmos pueden detectar automáticamente objetos como vehículos, áreas de construcción, carreteras, cultivos y tipos de techos (segmentación de imágenes). Ambos enfoques son cruciales ya que estas características u objetos presentan una correlación significativa con la riqueza y los ingresos locales.

El incentivo principal para desarrollar un modelo de inteligencia artificial que permita predecir ingresos geográficos a partir de imágenes satelitales radica en que, en muchas ocasiones, es posible distinguir visualmente entre zonas de ingresos alto, medio y bajo sin necesidad de información adicional. La meta es establecer un modelo que capture esta función o mapeo entre atributos observables en las imágenes, como la forma de las construcciones, los materiales utilizados, las superficies asfaltadas o la presencia de áreas verdes, y los ingresos promedio de los residentes en esas áreas.

El desarrollo de un modelo capaz de predecir ingresos promedio basándose únicamente en imágenes satelitales del área de interés conllevaría múltiples aplicaciones de gran relevancia. En primer lugar, contribuiría al campo de la estimación de áreas pequeñas al mejorar tanto la resolución espacial, al ofrecer predicciones sin limitación por la desagregación censal, como la resolución temporal, al permitir la predicción de ingresos usando imágenes satelitales en distintos momentos que no sean los de los censos. En segundo lugar, un modelo generalizado para diversas ciudades, más allá del área metropo-

litana, podría llenar el vacío de información en zonas sin datos disponibles empleando solo imágenes satelitales. En tercer lugar, los mapas generados mediante esta técnica podrían facilitar análisis de impacto localizados, como la evaluación de los efectos de nuevas infraestructuras, al comparar las predicciones del modelo antes y después de un cambio y seleccionar grupos de control adecuados. En cuarto lugar, la mejora en resolución espacial y temporal de los mapas de ingresos podría mejorar la evaluación de la focalización de programas sociales, ya que se cuenta con información geolocalizada de los beneficiarios. Estas estimaciones prescindirían de datos censales potencialmente desactualizados o encuestas, permitiendo la evaluación de la distribución de ingresos estimados según la ubicación residencial. Finalmente, estas técnicas podrían ayudar a desarrollar mejores estrategias de focalización de programas sociales, permitiendo nuevos criterios de exclusión.

A lo largo de esta sección se prestan algunos avances en torno a la generación de estimaciones espaciales del ingreso utilizando modelos de inteligencia artificial para predecir el ingreso a partir de imágenes satelitales. Estos desarrollos, que presentan una amplia variedad de aplicaciones, buscan reflejar la capacidad que esta tecnología podría tener en un futuro cercano. El mayor limitante a este tipo de metodologías, actualmente, es tecnológico, ya que la implementación de estos modelos requiere entrenar por varios días los modelos de inteligencia artificial, y este tiempo crece en relación al tamaño de la base de datos. Por ello, los análisis que siguen se limitan únicamente al AMBA y no abarcan a la totalidad de la Argentina.

Figura 8: Imágenes Satelitales de muy alta resolución (Google Maps)



(a) Barrio Los Ombúes, Tigre.



(b) Barrio San Justo, La Matanza.



(c) Barrio 31, Ciudad de Buenos Aires.

4.1. Metodología

Para actualizar los mapas generados con los datos del Censo 2010, se utilizarán imágenes satelitales de alta resolución obtenidas de la aplicación Google Maps para el AMBA. Se construyó una base de datos con una imagen representativa de cada radio censal. Para ello, se seleccionó aleatoriamente un punto dentro de cada radio censal y se capturó una imagen de $150m \times 150m$ centrada en ese punto. En muchos radios censales urbanos esta imagen contemplará la totalidad del radio, mientras que en áreas de menor densidad poblacional esta imagen será un subconjunto aleatorio del radio.

Habiendo construido la base de imágenes, se entrenó una Red Neuronal Convolutiva para que extraiga características relevantes de las imágenes que se correlacionen con el ingreso estructural por área geográfica. Las redes neuronales convolucionales son un tipo de modelo de aprendizaje profundo que se utiliza para procesar y analizar datos de imágenes. Las CNN se basan en la arquitectura de red neuronal artificial, pero tienen capas especiales de convolución diseñadas específicamente para procesar datos de imagen. Como arquitectura base de la CNN se utilizará la familia de modelos EfficientNetV2 desarrollados por Tan and Le (2021). Estos modelos son capaces de lograr un alto rendimiento con una cantidad relativamente baja de parámetros. Además, plantean un método de escalabilidad del modelo que se basa en aumentar de forma conjunta la resolución de las imágenes provistas, la cantidad de capas del modelo y la cantidad de canales que contiene cada capa. Se utilizaron los parámetros de ImageNet disponibles en Keras, y se entrenó únicamente la última capa densa para calibrar el modelo preentrenado a las imágenes utilizadas. Por lo tanto, el modelo es fácilmente adaptable al equipamiento disponible y podrá ser mejorado según la disponibilidad de *hardware* para alcanzar una mayor precisión en la estimación. Además, este tipo de modelos tiene capacidad de aplicación generalizada, es decir, que un modelo de inteligencia artificial como el planteado podría entrenarse con imágenes de varios países, y generar predicciones de ingreso espacial incluso en lugares donde no existe información disponible.

4.2. Resultados

Para evaluar la performance de un modelo de aprendizaje profundo como el utilizado, usualmente se divide aleatoriamente el conjunto de datos —las imágenes de Google Maps en conjunto con las etiquetas del ingreso estructural de cada radio censal— en dos partes: un conjunto de entrenamiento y un conjunto de prueba. El conjunto de entrenamiento, que en este caso es del 80% de las imágenes, se destina a entrenar propiamente el modelo, mientras que el restante 20% son imágenes que se utilizarán para validar si el modelo funciona como se espera, ya que son imágenes que nunca habrá visto. Si el ajuste del conjunto de prueba es apropiado, se considera que el modelo alcanzó la capacidad de generalizar el aprendizaje alcanzado y por lo tanto resulta útil para la aplicación de interés.

La Figura 9 muestra la correlación para el conjunto de prueba entre los valores predichos por el modelo y los valores “reales” del ingreso estructural de cada radio censal. Como puede observarse, la relación entre ambas variables alcanza un R^2 de 0.77, un resultado muy favorable para el modelo. A su vez, se observa que, si bien existe una dispersión considerable entre ambos valores, las predicciones siguen una tendencia central muy clara en torno a la diagonal principal, al tiempo que no se observan grandes sobreestimaciones en los deciles más bajos ni subestimaciones en los más altos, hechos que invalidarían muchas de las aplicaciones de este tipo de mapas, como el análisis de la focalización de programas sociales. A pesar de esto, sí se observa un leve sesgo de las predicciones hacia la media, que se refleja en la pendiente de la recta de regresión menor a 1.

Una vez validado el modelo, es posible utilizar los parámetros del mismo para generar predicciones sobre toda el Área Metropolitana de Buenos Aires, con lo cual se reconstruye un nuevo mapa, basado

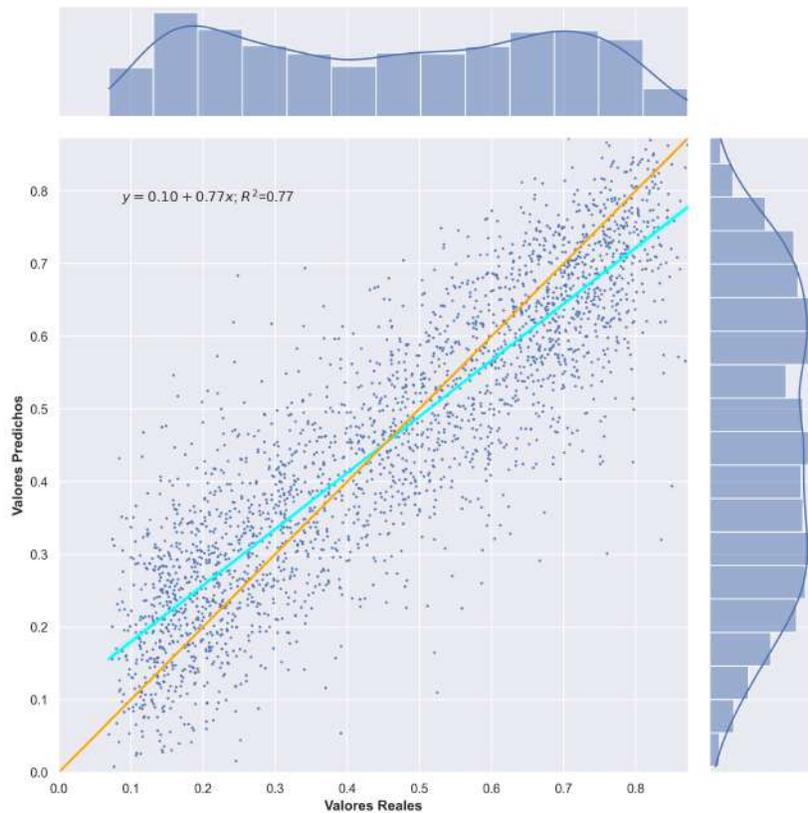


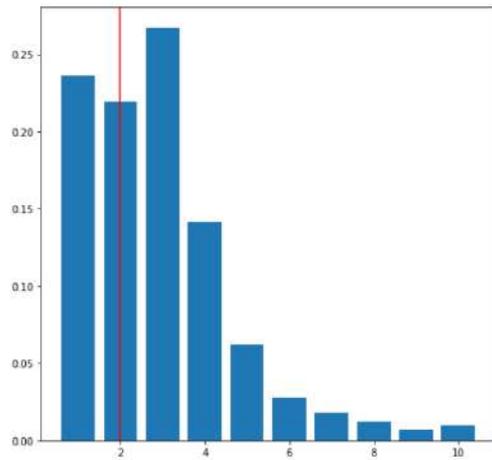
Figura 9: Comparación entre el ingreso estructural y las predicciones del grupo de prueba

enteramente en las imágenes y sin utilizar información censal. Este mapa se construye asignando una imagen para cada radio censal, y consecuentemente generando la predicción del modelo para tales imágenes.

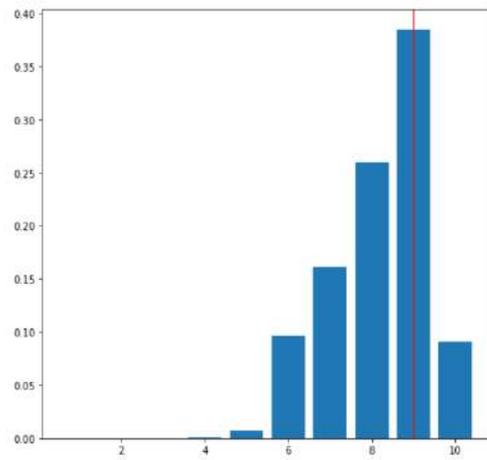
La Figura 10 muestra algunos ejemplos de estas predicciones: para cada imagen (lado derecho de la figura) se presenta con una línea roja vertical el decil del ingreso estructural correspondiente a su radio censal. Por otra parte, las columnas azules representan la probabilidad asignada por el modelo de que esa imagen se corresponda con cada decil. A cada imagen se le asigna como predicción el decil que tenga una mayor probabilidad, es decir, el decil correspondiente a la columna más alta. Entre los ejemplos, es importante destacar que tanto el caso (a) como el (c) son predicciones “incorrectas” ya que la predicción no se corresponde con el valor real del ingreso estructural —la columna más alta no coincide con la recta vertical roja. A pesar de eso, es claro que la distribución de probabilidades se encuentra cercana al verdadero valor, lo que indica que si bien el modelo presenta cierto error, este es bajo y no genera grandes errores predictivos —por ejemplo, identificar un barrio cerrado con uno del primer quintil del ingreso.

Utilizando estos insumos, la Figura 11 muestra las predicciones para una imagen aleatoria de cada radio censal del AMBA. Como puede observarse, estos resultados son cualitativamente similares a los

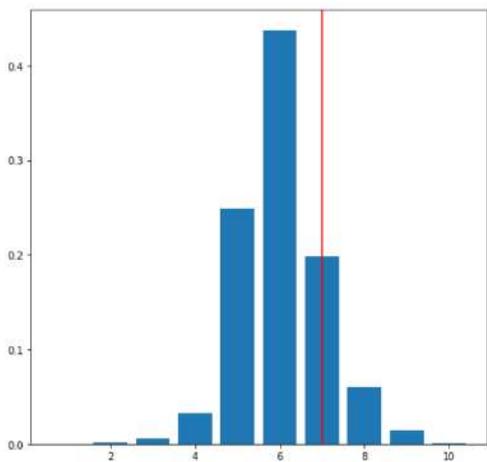
Figura 10: Distribución de predicciones por decil del ingreso estructural, imágenes seleccionadas



(a) Ingreso bajo



(b) Ingreso alto



(c) Ingreso medio

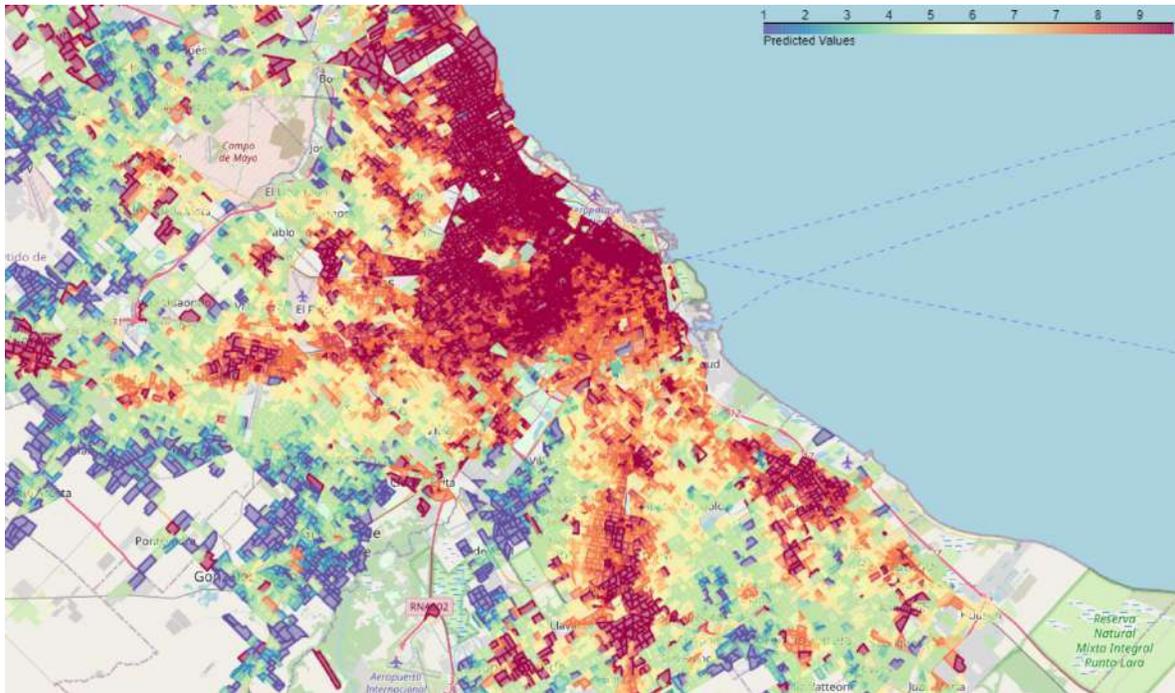
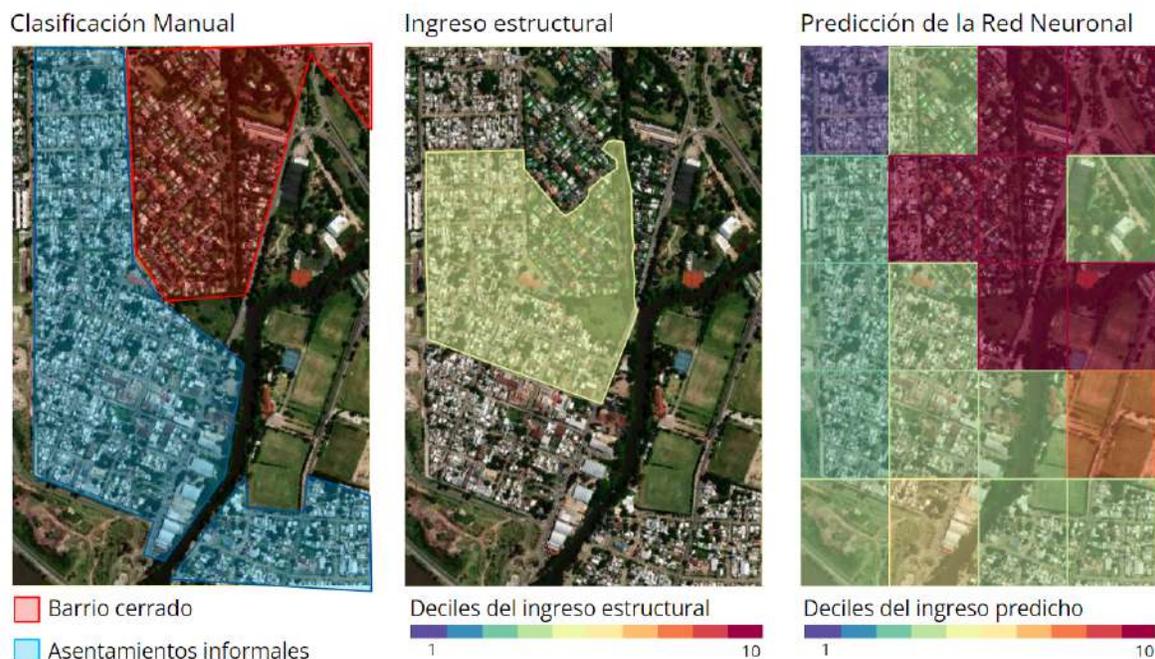


Figura 11: Mapa de predicciones del modelo EfficientNetV2-L (Deciles)

mapas originales del ingreso estructural por radio censal, pero surgen de estimaciones que sólo utilizan las imágenes satelitales. Este método presenta dos ventajas claras. Por un lado, su aplicación es posible a áreas y momentos de donde no se encuentran datos censales. En este caso, las imágenes satelitales se corresponden a 2018, por lo que los resultados pueden interpretarse como una actualización de las estimaciones del Censo 2010. Por otra parte, dado que las imágenes satelitales utilizadas deben ser de $150mts \times 150mts$, las predicciones pueden condicionarse a esas imágenes, lo que permitirá ofrecer una mayor granularidad, rompiendo la estructura de los radios censales.

Un ejemplo de aplicación reducido de esta mejora de la granularidad se presenta en la Figura 12. En ella se presentan, a la izquierda, la imagen satelital original con las características espaciales que presenta: se trata de una imagen de Tigre, una zona que tiene zonas contiguas entre barrios cerrados —típicamente de personas con ingresos muy altos— y asentamientos informales —típicamente de personas con ingresos muy bajos. En la figura central se muestra la forma del radio censal de esa imagen, donde se ve que el mismo atraviesa tanto un asentamiento informal como parte del barrio cerrado. El ingreso estructural, como es de esperarse, se corresponde al decil 5, es decir, en promedio, las personas tienen un “ingreso medio”. Lógicamente, esta media se encuentra lejos de ser representativa del radio censal ya que existe una diferencia muy grande entre las personas que viven a un lado y al otro del radio. La figura de la derecha muestra las predicciones del modelo para una grilla de imágenes del radio censal. En este caso, las predicciones identifican de forma mucho más clara el salto discreto en los ingresos: a la derecha de la imagen, los hogares parecen pertenecer al decil 10, mientras que a la izquierda pertenecen a deciles bajos. Este tipo de aplicaciones permite generar un análisis mucho

Figura 12: Comparación de predicciones de área pequeña



más preciso y pormenorizado del ingreso, facilitando estudios de focalización, evaluaciones de impacto y análisis intertemporales de urbanización.

5. Conclusiones

Este trabajo presenta la primera estimación del ingreso estructural con un alto grado de desagregación geográfica. A lo largo del estudio, se han empleado todas las encuestas EPH semestrales desde 2003 hasta 2022, además de los últimos dos censos disponibles (2001 y 2010). Los resultados de esta estimación muestran la validez del modelo al comparar las proyecciones con los valores reales. Además, se evidencia una alta correlación entre el Ingreso estructural y otros indicadores de bienestar.

A partir de esta investigación no solo se proporciona una nueva herramienta que permite comprender las dinámicas de ingresos a niveles geográficos más precisos, sino que también se allana el camino hacia nuevas posibilidades de exploración. Con relación al modelo, en primer lugar, se podrían explorar especificaciones alternativas a mínimos cuadrados ordinarios. En segundo lugar, resultaría deseable extrapolar el desarrollo sociodemográfico entre los censos y, así, obtener una estimación de las covariables para cada año.

Además, el potencial de integrar imágenes satelitales abre una nueva perspectiva en la estimación del

bienestar de los individuos, permitiendo la incorporación de datos espaciales adicionales que pueden contribuir a extender la aplicabilidad de esta metodología a regiones con imágenes satelitales disponibles.

Referencias

- Beccaria, L. and Gluzmann, P. A. (2013), Medición de los ingresos y la pobreza oficial en américa latina y el caribe, Technical report, Documento de Trabajo.
- Blumenstock, J., Cadamuro, G. and On, R. (2015), ‘Predicting poverty and wealth from mobile phone metadata’, *Science* **350**, 1073–1076.
- Burke, M., Driscoll, A., Lobell, D. B. and Ermon, S. (2021), ‘Using satellite imagery to understand and promote sustainable development’, *Science* **371**.
- Ciaschi, M. (2021), ‘Análisis distributivo utilizando información satelital. el caso de argentina’, *Estudios económicos* **38**, 5–38.
- Cuong, N. V., Truong, T. N. and van der Weide, R. (2010), ‘Poverty and inequality maps in rural vietnam: An application of small area estimation’, *Asian Economic Journal* **24**, 355–390.
- Department of Census and Statistics and World Bank (2015), ‘The spatial distribution of poverty in sri lanka department of census and statistics-sri lanka poverty global practice, world bank group’.
URL: http://www.statistics.gov.lk/national_accounts/Press%20Release/2014ANNUAL.pdf
- Ebers, C., Lanjouw, P. and Leite, P. G. (2008), *Brazil Within Brazil: Testing The Poverty Map Methodology In Minas Gerais*, The World Bank.
- Elbers, C., Lanjouw, J. O. and Lanjouw, P. (2003), ‘Micro-level estimation of poverty and inequality’, *Econometrica* **71**, 355–364.
- Engstrom, R., Hersh, J. and Newhouse, D. (2022), ‘Poverty from space: Using high resolution satellite imagery for estimating economic well-being’, *The World Bank Economic Review* **36**, 382–412.
URL: http://www.statistics.gov.lk/poverty/SpatialDistributionOfPoverty2012_3.pdf
- Gasparini, L. C., Gluzmann, P. A. and Tornarolli, L. (2022), ‘Caracterización de la población vulnerable: una propuesta con estimaciones para argentina’, *Económica* **68**.
- Gluzmann, P. and Panigo, D. (2015), ‘Global search regression: A new automatic model-selection technique for cross-section, time-series, and panel-data regressions’, *The Stata Journal* **15**(2), 325–349.
- Henderson, J. V., Storeygard, A. and Weil, D. N. (2012), ‘Measuring economic growth from outer space’, *American Economic Review* **102**, 994–1028.
- Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B. and Ermon, S. (2016), ‘Combining satellite imagery and machine learning to predict poverty’, *Science* **353**, 790–794.
- Molina, I. and Rao, J. N. K. (2010), ‘Small area estimation of poverty indicators’, *Canadian Journal of Statistics* **38**, 369–385.

- Panigo, D., Gluzmann, P., Mocskos, E., Ungaro, A. M., Mari, V. and Monzón, N. (2020), ‘Globalsearchregression. jl: building bridges between machine learning and econometrics in fat-data scenarios’, **2**(13), 53.
- Panigo, D., Ungaro, A. M., Monzón, N., Tcach, A., Menduiña, J., Gluzmann, P., Barcellandi, A. R. and Panigo, N. (2023), ‘Accelerating economic research with julia’, *presentado en JuliaCon 2023* .
- Sheehan, E., Meng, C., Tan, M., Uzkent, B., Jean, N., Lobell, D., Burke, M. and Ermon, S. (2019), ‘Predicting economic development using geolocated wikipedia articles’.
- Steele, J. E., Sundsøy, P. R., Pezzulo, C., Alegana, V. A., Bird, T. J., Blumenstock, J., Bjelland, J., Engø-Monsen, K., de Montjoye, Y.-A., Iqbal, A. M., Hadiuzzaman, K. N., Lu, X., Wetter, E., Tatem, A. J. and Bengtsson, L. (2017), ‘Mapping poverty using mobile phone and satellite data’, *Journal of The Royal Society Interface* **14**, 20160690.
- Tan, M. and Le, Q. V. (2021), ‘Efficientnetv2: Smaller models and faster training’.
URL: <http://arxiv.org/abs/2104.00298>
- Tibshirani, R. (1996), ‘Regression shrinkage and selection via the lasso’, *Journal of the Royal Statistical Society. Series B (Methodological)* **58**(1), 267–288.
URL: <http://www.jstor.org/stable/2346178>
- Tornarolli, L. (2018), Series comparables de indigencia y pobreza: una propuesta metodológica, Technical report, Documento de Trabajo.
- Weber, I., Kashyap, R. and Zagheni, E. (2018), ‘Using advertising audience estimates to improve global development statistics’.
URL: <https://www.itu.int/en/journal/002/Pages/default.aspx>
- Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., Ermon, S. and Burke, M. (2020), ‘Using publicly available satellite imagery and deep learning to understand economic well-being in africa’, *Nature Communications* **11**, 2583.