

Predicciones agregadas de pobreza con información a escala micro y macro

Evaluación, diagnóstico
y propuestas

Walter Sosa Escudero
Magdalena Cornejo



NACIONES UNIDAS

CEPAL

Gracias por su interés en esta publicación de la CEPAL



Si desea recibir información oportuna sobre nuestros productos editoriales y actividades, le invitamos a registrarse. Podrá definir sus áreas de interés y acceder a nuestros productos en otros formatos.

 www.cepal.org/es/publications

 www.cepal.org/apps

SERIE

ESTUDIOS ESTADÍSTICOS

103

Predicciones agregadas de pobreza con información a escala micro y macro

Evaluación, diagnóstico y propuestas

Walter Sosa Escudero
Magdalena Cornejo



NACIONES UNIDAS

CEPAL

Este documento fue preparado por Walter Sosa Escudero y Magdalena Cornejo, Consultores de la Unidad de Estadísticas Sociales de la División de Estadísticas de la Comisión Económica para América Latina y el Caribe (CEPAL), bajo la coordinación de Xavier Mancero, Jefe, y Álvaro Fuentes, Estadístico de dicha Unidad, en el marco de las actividades del proyecto "Strengthening social protection for pandemic responses: identifying the vulnerable, aiding recovery and building resilience" de la Cuenta de las Naciones Unidas para el Desarrollo.

Las opiniones expresadas en este documento, que no ha sido sometido a revisión editorial, son de exclusiva responsabilidad de los autores y pueden no coincidir con las de la Organización o las de los países que representa.

Publicación de las Naciones Unidas
ISSN: 1680-8789 (versión electrónica)
ISSN: 1994-7364 (versión impresa)
LC/TS.2022/95
Distribución: L
Copyright © Naciones Unidas, 2022
Todos los derechos reservados
Impreso en Naciones Unidas, Santiago
S.22-00372

Esta publicación debe citarse como: W. Sosa Escudero y M. Cornejo, "Predicciones agregadas de pobreza con información a escala micro y macro: evaluación, diagnóstico y propuestas", *serie Estudios Estadísticos*, N° 103 (LC/TS.2022/95), Santiago, Comisión Económica para América Latina y el Caribe (CEPAL), 2022.

La autorización para reproducir total o parcialmente esta obra debe solicitarse a la Comisión Económica para América Latina y el Caribe (CEPAL), División de Documentos y Publicaciones, publicaciones.cepal@un.org. Los Estados Miembros de las Naciones Unidas y sus instituciones gubernamentales pueden reproducir esta obra sin autorización previa. Solo se les solicita que mencionen la fuente e informen a la CEPAL de tal reproducción.

Índice

Resumen	5
Introducción	7
I. Enfoques recientes en predicciones de la pobreza: nowcasting y aprendizaje	9
II. Fuentes de datos para la predicción	13
III. El enfoque micro-macro	17
IV. Resultados y desempeño del modelo base en 2019.....	19
V. Extensiones del modelo base y resultados para 2019	23
A. Efectos asimétricos	27
B. Incorporación de información externa.....	29
C. Pronósticos de pobreza para los países faltantes.....	31
VI. Evaluación del desempeño predictivo para el período 2003 a 2019.....	33
A. Pronósticos de pobreza en mujeres	40
B. Pronósticos de pobreza en desocupados	40
C. Pronósticos de pobreza en jóvenes.....	49
VII. Recomendaciones para la práctica	55
Comentarios finales	59
Bibliografía.....	61
Anexos	63
Anexo 1	64
Anexo 2	70
Serie Estudios Estadísticos: números publicados.....	73

Cuadros

Cuadro 1	Variables relevadas.....	14
Cuadro 2	Valores perdidos por país y variable.....	14
Cuadro 3	Valores perdidos por país y variable luego de la interpolación.....	15
Cuadro 4	Proyecciones agregadas de pobreza y pobreza extrema para 2019.....	19
Cuadro 5	Principales medidas de desempeño de pronóstico.....	20
Cuadro 6	Evaluación desagregada de las proyecciones de 2019.....	21
Cuadro 7	Pronósticos para 2019 de crecimiento ($\Delta \ln p_{i,t}$) y desigualdad ($G_{i,t}$).....	25
Cuadro 8	Comparación de los valores pronosticados de crecimiento (β_{2019}) y cambio en la desigualdad (α_{2019}).....	26
Cuadro 9	Comparación de pronósticos de las tasas de pobreza para 2019.....	28
Cuadro 10	Comparación de los pronósticos de β_{2019} y α_{2019} incorporando variables externas.....	30
Cuadro 11	Pronósticos para 2019 de β y α para los países faltantes.....	31
Cuadro 12	Pronósticos de tasas de pobreza de 2019 para los países faltantes.....	31
Cuadro 13	Error medio de pronóstico de los distintos modelos (2003-2019).....	35
Cuadro 14	La raíz del error cuadrático medio de pronóstico de los distintos modelos (2003-2019).....	36
Cuadro 15	El error medio absoluto de pronóstico de los distintos modelos (2003-2019).....	37
Cuadro 16	El error mediano absoluto del pronóstico de los distintos modelos (2003-2019).....	39
Cuadro 17	El error medio de pronóstico de pobreza en mujeres.....	41
Cuadro 18	Raíz del error cuadrático medio del pronóstico de pobreza en mujeres.....	42
Cuadro 19	Error medio absoluto del pronóstico de pobreza en mujeres.....	43
Cuadro 20	Error mediano absoluto del pronóstico de pobreza en mujeres.....	44
Cuadro 21	El error medio de pronóstico de pobreza en desocupados.....	45
Cuadro 22	Raíz del error cuadrático medio del pronóstico de pobreza en desocupados.....	46
Cuadro 23	Error medio absoluto del pronóstico de pobreza en desocupados.....	47
Cuadro 24	Error mediano absoluto del pronóstico de pobreza en desocupados.....	48
Cuadro 25	El error medio de pronóstico de pobreza en jóvenes.....	50
Cuadro 26	Raíz del error cuadrático medio del pronóstico de pobreza en jóvenes.....	51
Cuadro 27	Error medio absoluto del pronóstico de pobreza en jóvenes.....	52
Cuadro 28	Error mediano absoluto del pronóstico de pobreza en jóvenes.....	53
Cuadro A1	Argentina.....	64
Cuadro A2	Bolivia.....	64
Cuadro A3	Brasil.....	65
Cuadro A4	Costa Rica.....	65
Cuadro A5	Ecuador.....	66
Cuadro A6	El Salvador.....	66
Cuadro A7	Honduras.....	67
Cuadro A8	Panamá.....	67
Cuadro A9	Paraguay.....	68
Cuadro A10	Perú.....	68
Cuadro A11	República Dominicana.....	69
Cuadro A12	Uruguay.....	69

Gráficos

Gráfico 1	Evolución temporal del PIB per cápita, medido en dólares constantes de 2010.....	16
Gráfico 2	Evolución temporal del Coeficiente de Gini.....	16
Gráfico A1	Evolución del ingreso medio per cápita (en moneda nacional).....	70
Gráfico A2	Evolución de la tasa de pobreza.....	71
Gráfico A3	Evolución de la tasa de pobreza extrema.....	72

Resumen

En este documento se discuten y revisan diversas alternativas para realizar pronósticos de pobreza para varios países de América Latina. El punto de partida es el modelo base desarrollado por CEPAL y luego se generan variantes que exploran estrategias novedosas asociadas a las técnicas de *machine learning* (aprendizaje automático). Se parte de la construcción de un panel para 12 países de la región entre 2000 y 2019 y se realiza un análisis comparativo de las proyecciones realizadas de las tasas agregadas de pobreza y pobreza extrema. Se evalúan distintas alternativas de pronóstico de pobreza que buscan explotar la naturaleza micro-macro de los datos, la dinámica temporal de las series, la heterogeneidad del panel y el uso de técnicas de *machine learning* que permiten lidiar con la complejidad de los modelos. El desempeño predictivo fue evaluado tanto a nivel agregado como a través de grupos de individuos (i.e. mujeres, desocupados y jóvenes).

Introducción

Difícilmente existan sectores del conocimiento ajenos a la revolución de los métodos de aprendizaje automático o *machine learning*, y la economía claramente no es una excepción. Así y todo, en comparación con casi todas las disciplinas científicas, la economía fue prudente en la incorporación de dichos métodos, posiblemente debido a la prevalencia de métodos causales que imperaron la práctica empírica de la profesión durante los últimos 30 años. La principal diferencia entre los métodos econométricos estándar y los de *machine learning* radica en que los primeros tienden a tomar a los modelos como dados (por una teoría o experiencia previa) y concentran esfuerzos en estimar parámetros o aspectos de interés (usualmente efectos causales) de la mejor manera posible, permitiendo a la vez realizar inferencias válidas, es decir, mensurar la incertidumbre de modo que se pueda construir intervalos de confianza o evaluar hipótesis relevantes. Por el contrario, los métodos de *machine learning* se focalizan en realizar predicciones confiables, para lo cual el modelo subyacente no es tomado como fijo o dado, sino que es construido o alterado (“entrenado, aprendido”), en forma iterativa, con el objetivo de realizar mejores predicciones fuera de la muestra disponible. Ver Sosa Escudero (2019) para más detalles sobre la distinción entre los métodos tradicionales y los de *machine learning*.

Consecuentemente, las ideas de aprendizaje automático hallaron un terreno fértil en el ámbito de las predicciones y la búsqueda de patrones en los datos, que contrasta con la necesidad de estimar efectos causales que gobernó a muchas ramas de la economía. Pese a ello, la relevancia de predecir y medir fenómenos importantes, como la pobreza, los tipos de cambio o el nivel de actividad económica, son aspectos tan relevantes como los de indagar en relaciones de causa-efecto, toda vez que la política pública requiere tanto de aspectos predictivos como causales, como señala el iluminador artículo de Kleinberg et al. (2015), que muestra que ambas dimensiones son más bien complementos que sustitutos.

Este trabajo apela a algunas estrategias novedosas de *machine learning* para realizar pronósticos de pobreza para varios países de América Latina, en un contexto concreto. La CEPAL utiliza una metodología basada en un enfoque “micro – macro”, que combina información agregada (macro) con datos de encuestas de hogares (micro), cuyo funcionamiento se describe con detalle en el capítulo III y en las referencias ahí citadas. El resultado de esta metodología es una predicción de la tasa de pobreza agregada para un periodo

y en un país en particular. Los insumos para la predicción son: 1) microdatos de ingresos para distintos periodos y países, 2) la tasa de variación del ingreso medio, para cada país y período, 3) las tasas de variación del coeficiente de Gini, para cada país y periodo. El así llamado “modelo base” produce una predicción para la tasa de pobreza de un país y periodo, en base a la información disponible descripta anteriormente correspondiente al periodo inmediatamente anterior al de la predicción.

Este trabajo realiza varias tareas. En primer lugar, realiza una evaluación de la capacidad predictiva del “modelo base”, utilizando varias métricas que son estándar en la literatura de *machine learning*, derivadas de la “matriz de confusión” y sus derivaciones (exactitud, especificidad, etc.). Luego se exploran varios modelos alternativos, siempre utilizando la misma información del modelo base, de modo que la contribución es esencialmente técnica, es decir, no se explora en esta etapa usar información alternativa o adicional. La exploración utiliza varias estrategias novedosas de regularización, incluyendo algunas que han cobrado relevancia dentro del paradigma reciente de *machine learning*.

A su vez, se evalúan las distintas alternativas de pronóstico de pobreza sugeridas desde tres perspectivas diferentes. En primer lugar, se evalúa el desempeño de los pronósticos de 2019 para un grupo de 24 “pseudo-individuos” con el objetivo de aumentar la representatividad de los distintos grupos. En segundo lugar, se realizan ejercicios adicionales sobre los modelos de pronósticos presentados evaluando en particular efectos de asimetría, incorporando información externa que pueda mejorar el poder predictivo de los modelos en la región y generando los pronósticos de pobreza de 2019 para los países faltantes del panel original. En tercer lugar, realizamos un ejercicio de validación cruzada de los distintos modelos. Es decir, evaluamos el desempeño de éstos entre 2003 y 2019 (17 años) para sacar conclusiones más generalizables y empíricamente testeables. En forma adicional, dicho ejercicio de validación cruzada es replicado para los pronósticos de pobreza y pobreza extrema de algunos grupos de interés: las mujeres, los desocupados y los jóvenes (entre 15 y 24 años).

Es importante mencionar que el objetivo de este trabajo es predecir la pobreza a nivel agregado, es decir, para distintos países y períodos. Si bien incipiente, existen varios trabajos recientes que implementan métodos de *machine learning* para predecir el estatus individual de pobreza. Ver Sosa Escudero et al. (2021) para una revisión reciente de este tipo de estudios. Estas alternativas, si bien interesantes y relevantes cuando ese es el objetivo, no necesariamente se traducen en mejoras cuando el objetivo es predecir la tasa de pobreza agregada.

La estructura del trabajo es la siguiente. El capítulo I revisa algunas ideas recientes del problema de la predicción de la pobreza (*nowcasting*). Posteriormente, el capítulo II discute con detalle las fuentes de información utilizada. El modelo base, “micro-macro” utilizado por CEPAL es discutido en el capítulo III, y en el capítulo IV se presentan los resultados empíricos del mismo. El capítulo V explora varias alternativas de este modelo, incorporando información dinámica en un contexto de selección de modelos. El capítulo VI realiza una evaluación exhaustiva de la performance predictiva de los modelos implementados, en varias dimensiones, es decir, tanto a nivel de país como de grupos específicos de la población (por género, estatus de empleo, etc.). Posteriormente, en el capítulo VII se describen algunas recomendaciones para la práctica.

I. Enfoques recientes en predicciones de la pobreza: *nowcasting* y aprendizaje

Las técnicas de predicción inmediata o *nowcasting* tienen sus orígenes en la meteorología y podríamos citar el trabajo de Giannone et al. (2008) como el trabajo pionero dentro del ámbito de la economía que desarrolló los métodos de pronóstico en tiempo real sobre datos macroeconómicos (i.e. la tasa de crecimiento del producto bruto interno).

En años recientes, las técnicas de *nowcasting* comenzaron a utilizarse para monitorear y estimar la tasa de pobreza en tiempo real valiéndose de distintas variables socioeconómicas publicadas en una mayor frecuencia.

En línea con el enfoque utilizado por CEPAL, Aguilar et al. (2019) discuten cómo grandes conjuntos de información (i.e. provenientes de las bases de datos *World Economic Outlook* y *World Development Indicators*) junto con el uso de técnicas de aprendizaje estadístico (técnicas de regularización, *bosques aleatorios* (“*random forests*”) y *potenciación del gradiente* (“*gradient boosting*”)) pueden mejorar la precisión de los pronósticos actuales de pobreza que realiza el Banco Mundial. En este trabajo, siguen un enfoque similar al propuesto por CEPAL ya que utilizan estas técnicas y datos para predecir la tasa de crecimiento medio del ingreso y calculan las tasas de pobreza aplicando dichas tasas predichas de crecimiento por igual a todos los hogares según los últimos datos observados en las encuestas de microdatos. Los resultados de este trabajo sugieren que hay ganancias menores en el desempeño de los pronósticos, pero advierten que podrían obtenerse mejoras más significativas si se predice no solo la tasa de crecimiento del ingreso sino del coeficiente de Gini.

Respecto a los trabajos más recientes tendientes a predecir la pobreza en el contexto de la crisis del COVID-19 sobre la pobreza, Brum y De Rosa (2021) proponen un enfoque que busca evaluar el impacto de la crisis sobre los ingresos basado en microsimulaciones (usando microdatos a nivel de los hogares) combinado con estimaciones sobre la viabilidad de trabajar desde el hogar e información sobre algunas políticas públicas puntuales (e.g. como transferencias de ingreso o

seguros de desempleo) y pronósticos de contracción del PIB. La metodología es ilustrada a partir de la aplicación a los datos del segundo trimestre de 2020 en Uruguay. Las estimaciones de pobreza surgen a partir de la comparación de los ingresos proyectados con respecto de las líneas de pobreza.

La falta de información suele ser todo un desafío para los estudios de ingreso y bienestar. Varios trabajos combinan distintas fuentes de información para mejorar la disponibilidad de datos, la frecuencia y la granularidad. Elbers et al. (2003) proponen un procedimiento estadístico que combina encuestas y datos censales. Los autores son capaces de estimar la pobreza para poblaciones tan pequeñas como de 15.000 hogares. Muchos trabajos posteriores que aplican técnicas de *machine learning* en pobreza y desigualdad están relacionados con este trabajo (Blumenstock et al., 2015, Jean et al., 2016).

Las técnicas de aprendizaje no supervisado, tales como el agrupamiento (*clustering*) y las distintas estrategias de reducción de la dimensionalidad, permiten abordar el problema de la predicción de la pobreza desde su naturaleza multidimensional. En esta línea, el trabajo de Caruso et al. (2015) propone una metodología en dos etapas en la que primero identifican los grupos pobres usando métodos de *clustering* basados en un conjunto amplio de atributos, y segundo, reducen la dimensión del bienestar original encontrando el conjunto de atributos más pequeño que reproduce fehacientemente la clasificación de pobres obtenida en la primera etapa.

Si el problema de dimensionalidad se reduce a una única dimensión, estas técnicas son útiles también para construir índices. Edo et al. (2021) proponen un enfoque novedoso en el que reducen la dimensionalidad del bienestar para identificar la clase media.

Lucchetti (2018), por su parte, utiliza métodos supervisados de *machine learning* para estimar la dinámica de la riqueza de los hogares ante la ausencia de datos de panel. La metodología propuesta consiste en estimar parámetros de un modelo de ingresos (logarítmicos) en la primera ronda de datos transversales usando LASSO¹, para predecir los ingresos de hogares encuestados en la segunda ronda, pero que falta su información de la primera ronda. El enfoque es validado comparando la dinámica de los ingresos estimados a partir de los datos de corte transversal con aquellos derivados de los datos en panel en Perú. El autor sugiere que utilizar un procedimiento de regularización como LASSO ante la ausencia de datos longitudinales acerca de individuos o hogares en dos o más momentos del tiempo genera resultados alentadores para estimar la movilidad económica.

Sampi y Jooste (2020) proponen utilizar como indicador líder el “Google Mobility Index”, construido a partir del reporte de movilidad que elabora Google tras la pandemia del COVID 2019, para pronosticar las tasas de crecimiento mensuales del índice de predicción para una selección de países de América Latina y el Caribe. Utilizan datos en frecuencias mixtas para pronosticar dicha tasa de crecimiento de producción industrial entre enero de 2019 y abril de 2020 valiéndose tanto de los datos de Google para construir dicho índice de movilidad a partir de un análisis de componentes principales y extendiendo dicho índice hacia atrás usando información relacionada a la calidad de aire.

A través del uso de datos no tradicionales, como los teléfonos móviles y otros sensores digitales, Aiken et al. (2020) usan métodos supervisados de *machine learning* para identificar a hogares ultrapobres, respecto de otros hogares pobres no elegibles, que reúnen los requisitos para ser beneficiarios de un programa antipobreza en Afganistán. Los autores prueban que estas técnicas tienen un nivel de precisión equivalente a las medidas tradicionales de pobreza basadas en encuestas. Como antecedente al uso de este tipo de datos para estimar las características

¹ LASSO (*Least Absolute Shrinkage and Selection Operator*) fue introducido por Tibshirani (1996) y consiste en un método de análisis de regresión que realiza una selección de variables imponiendo restricciones sobre los coeficientes del modelo lineal.

socioeconómicas de los hogares podemos mencionar el trabajo de Blumenstock (2018). El principal objetivo de este trabajo es inferir las características de un individuo a partir de las “huellas digitales” que deja a través del uso de dispositivos digitales como teléfonos, redes sociales y otras tecnologías. Sin embargo, el autor advierte que estos modelos que son entrenados en un país en particular no pueden usarse para estimar las características en otro.

En esta línea, Jean et al. (2016) utilizaron imágenes satelitales para estimar, en una forma precisa, económica y escalable mediante el uso de redes neuronales, el gasto en consumo y riqueza de activos. En este trabajo complementan los datos de encuestas y satélites para cinco países africanos: Nigeria, Tanzania, Uganda, Malawi y Ruanda.

II. Fuentes de datos para la predicción

La CEPAL realiza sus proyecciones de pobreza utilizando los microdatos de las encuestas de hogares que se utilizan para medir pobreza de manera oficial en cada país (incluidas en el Banco de Datos de Encuestas de Hogares, BADEHOG) y dos variables principales: (1) el producto interno bruto (PIB) per cápita a precios constantes en dólares (base 2010) para proyectar la tasa de crecimiento del ingreso y (2) el coeficiente de Gini para proyectar la variación de la desigualdad en el ingreso.

Para este trabajo, se utilizaron esas mismas variables, complementadas con datos sobre el ingreso medio per cápita familia en moneda nacional, el coeficiente de Gini, las líneas de pobreza y pobreza extrema de las áreas urbanas, y las tasas de pobreza y pobreza extrema estimadas. Los datos del PIB per cápita provienen de CEPALSTAT, mientras que los datos del coeficiente de Gini provienen de CEPALSTAT y del SEDLAC² (base de datos socioeconómicos para América Latina y el Caribe desarrollada por el CEDLAS y el Banco Mundial). Para el caso del coeficiente de Gini se considera el dato nacional, salvo para Argentina durante todo el período y Uruguay hasta 2005 que se dispone del dato urbano.

El cuadro 1 resume las variables que fueron relevadas priorizando trabajar con datos publicados por CEPAL y complementados por estadísticas oficiales o de la SEDLAC.

² <https://www.cedlas.econo.unlp.edu.ar/wp/estadisticas/sedlac/>.

Cuadro 1
VARIABLES RELEVADAS

Variable	Fuente
Ingreso medio p/c familiar, en moneda nacional	SEDLAC (CEDLAS y Banco Mundial)
PIB p/c a precios constantes en dólares	CEPALSTAT
Coefficiente de Gini	CEPALSTAT y SEDLAC
Línea de pobreza urbana, en moneda nacional	CEPALSTAT y estadísticas oficiales
Línea de indigencia urbana, en moneda nacional	CEPALSTAT y estadísticas oficiales
Tasa de pobreza	CEPALSTAT
Tasa de pobreza extrema	CEPALSTAT

Fuente: Elaboración propia.

Con esta información, se construyó un panel para 18 países de la región a lo largo de los últimos 20 años (2000-2019). Del total de 360 observaciones posibles, el panel resultante logra una cobertura de solo el 70%, debido a que no se dispone de información para todos los años en todos los países. Los países con mayor proporción de valores perdidos fueron: Chile (60%), Guatemala (85%), Honduras (40%), México (50%), Nicaragua (80%) y Venezuela (65%)³.

La primera parte de los análisis realizados excluyen a los países con datos faltantes, lo que deja 12 países con datos consistentemente observables para 20 años. En este grupo la proporción de datos perdidos alcanza 13%, por lo que sigue siendo un panel levemente desbalanceado. El cuadro 2 muestra la distribución de valores perdidos por país y por variable a lo largo de 2000-2019.

Cuadro 2
Valores perdidos por país y variable

País	Ingreso medio	PIB p/c	Coef. de Gini	Línea pobreza	Línea indigencia	Pobreza	Pobreza extrema
Argentina	-	-	1	1	1	1	1
Bolivia (Estado Plurinacional de)	4	-	2	3	3	2	2
Brasil	3	-	2	3	3	2	2
Colombia	4	-	4	4	4	4	4
Costa Rica	-	-	-	-	-	-	-
Ecuador	3	-	1	4	4	3	3
El Salvador	1	-	-	3	3	2	2
Panamá	1	-	-	3	3	2	2
Paraguay	2	-	1	2	2	-	-
Perú	1	-	-	2	2	-	-
República Dominicana	1	-	-	1	1	-	-
Uruguay	1	-	-	2	2	1	1

Fuente: Elaboración propia.

³ Aun cuando Chile y México cuentan con encuestas regulares para la medición de los ingresos y la pobreza, estas no se llevan a cabo todos los años, sino cada 2 o 3 años. En Guatemala y Nicaragua se dispone de este tipo de encuestas solo para 3 años en el período considerado, y el dato más reciente es de 2014, al igual que en el caso de Venezuela.

Dado que estimaremos modelos dinámicos en panel y que queremos evitar el problema de la falta de datos y posibles discontinuidades que se generen, realizamos una interpolación de los datos usando *splines* cúbicos⁴. La naturaleza relativamente cíclica y suave de los valores de ingresos medios, pobreza y Gini sugiere que el uso de *splines* es una herramienta plausible. El cuadro 3 reporta cómo queda la distribución de valores perdidos por país y variable para nuestro período de análisis.

Cuadro 3
Valores perdidos por país y variable luego de la interpolación

País	Ingreso medio	PIB p/c	Coef. de Gini	Línea pobreza	Línea indigencia	Pobreza	Pobreza extrema
Argentina	-	-	1	1	1	1	1
Bolivia (Estado Plurinacional de)	1	-	-	1	1	-	-
Brasil	2	-	1	2	2	1	1
Colombia	2	-	2	2	2	2	2
Costa Rica	-	-	-	-	-	-	-
Ecuador	1	-	1	3	3	1	1
El Salvador	1	-	-	1	1	-	-
Panamá	1	-	-	1	1	-	-
Paraguay	2	-	1	2	2	-	-
Perú	1	-	-	2	2	-	-
República Dominicana	1	-	-	1	1	-	-
Uruguay	1	-	-	1	1	1	1

Fuente: Elaboración propia.

Si bien siguen existiendo valores faltantes, fundamentalmente al inicio y fin del período, el nivel de cobertura asciende al 92%. De todas formas, este leve desbalanceo será considerado en la estimación de varios de los modelos propuestos en este trabajo.

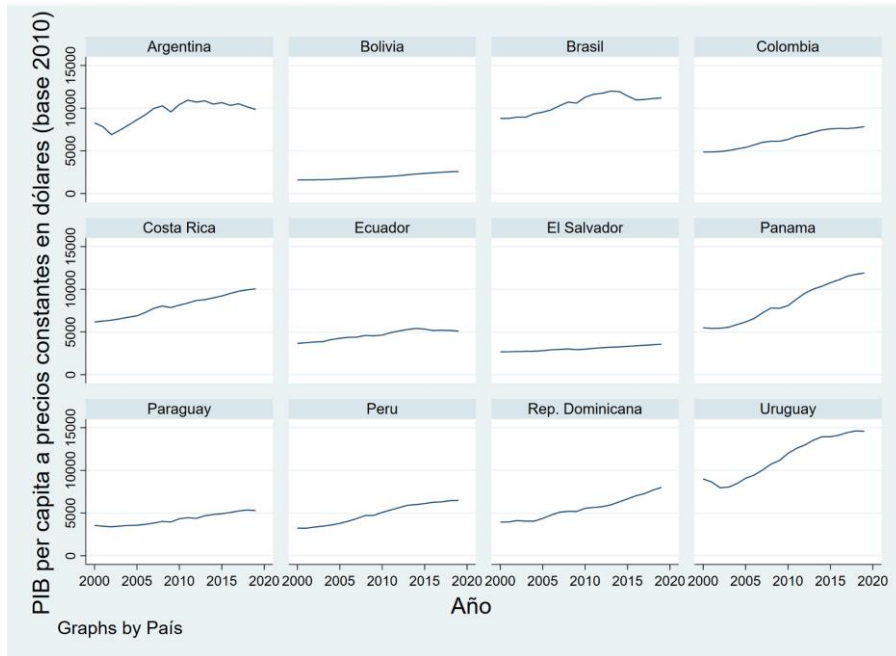
Los gráficos 1 y 2 muestran la evolución temporal del PIB per cápita (en dólares constantes de 2010) y del coeficiente de Gini para los 12 países analizados. El resto de los cuadros correspondientes a las otras variables relevadas pueden encontrarse en el anexo 2.

Como adelantásemos, se puede observar el comportamiento tendencial, aunque no siempre lineal, que muestra esta variable para todos los países. Cabe destacar que los modelos desarrollados utilizarán como variable dependiente la tasa de crecimiento anual del PIB per cápita real.

Respecto al coeficiente de Gini (ver gráfico 2), muestra una caída sistemática en el conjunto de países a lo largo del tiempo, indicando una mayor igualdad en la distribución del ingreso. Al igual que el PIB per cápita, esta variable muestra un comportamiento relativamente suave y persistencia en el tiempo, lo cual tendremos en consideración a la hora de desarrollar nuestros modelos de pronósticos.

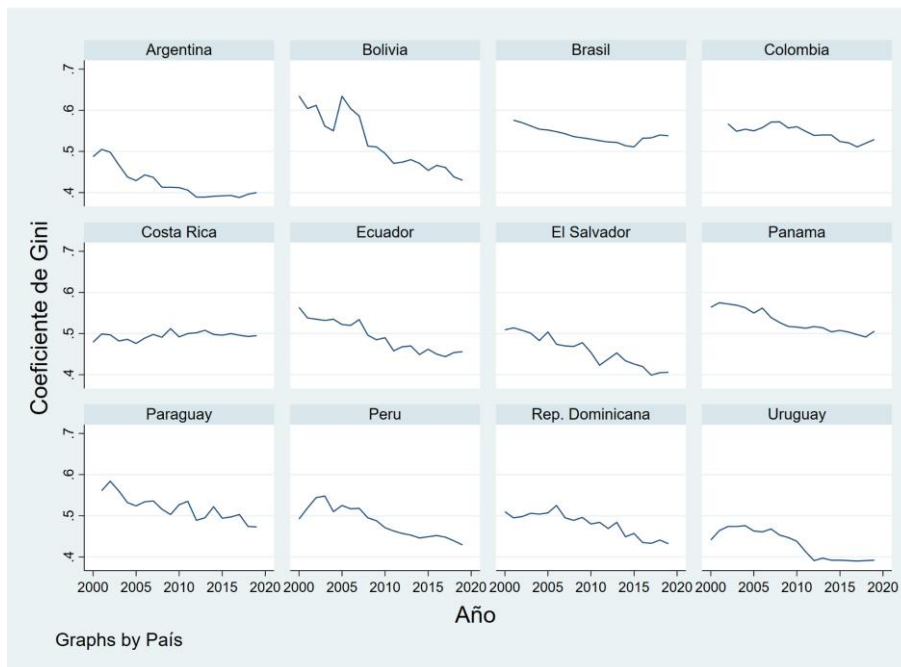
⁴ Los splines son funciones utilizadas para la interpolación de datos. El spline cúbico es una técnica de interpolación polinómica de grado 3 que conecta puntos de datos a través de una curva y es muy comúnmente empleado en la práctica ya que los polinomios resultantes son suaves y continuos.

Gráfico 1
Evolución temporal del PIB per cápita, medido en dólares constantes de 2010



Fuente: Elaboración propia.

Gráfico 2
Evolución temporal del Coeficiente de Gini



Fuente: Elaboración propia.

III. El enfoque micro-macro

Este capítulo describe el enfoque “micro-macro” usado habitualmente por CEPAL para realizar proyecciones de pobreza y que conforman el punto de partida de este trabajo. Ver CEPAL, IPEA y PNUD (2003) para más detalles.

El modelo base para la predicción proviene de la siguiente fórmula,

$$\widehat{y}_i^{t+1} = (1 + \beta)[(1 - \alpha)y_i^t + \alpha\mu^t]$$

donde:

- y_i^t es el ingreso por persona de cada hogar i en el año t ;
- μ^t es el ingreso medio por persona de todos los hogares en el año t ;
- α es un parámetro que expresa la reducción porcentual proyectada en el índice de Gini;
- β es el parámetro que da cuenta de la tasa de crecimiento del ingreso por persona, que se supone igual a la tasa de crecimiento proyectada por la CEPAL del PIB per cápita (en dólares constantes de 2010) para $t + 1$.

Los ingresos proyectados para $t + 1$ a partir de los datos de la encuesta de hogares en t permiten identificar a las personas pobres en $t + 1$ y calcular la respectiva tasa de pobreza. Dado que la proyección de ingresos está expresada en términos reales se mantiene constante el valor de la línea de pobreza del período t para proyectar pobreza en $t + 1$.

Entonces, las tasas de pobreza y de pobreza extrema quedan determinadas de la siguiente manera:

$$Tasa\ de\ pobreza = \frac{\sum_{i=1}^n I(\widehat{y}_i^{t+1} < lp^{t+1})}{n}$$

$$Tasa\ de\ pobreza\ extrema = \frac{\sum_{i=1}^n I(\widehat{y}_i^{t+1} < li^{t+1})}{n}$$

donde lp y li son las líneas de pobreza y pobreza extrema en $t + 1$ (expresadas en términos reales), respectivamente.

El parámetro α genera una modificación distributiva de los ingresos tomando un valor positivo en períodos de contracción económica y nulo en períodos de crecimiento económico. En años de contracción económica (cuando $\alpha > 0$), dicho deterioro se implementa en forma asimétrica aplicando la modificación distributiva sólo a aquellos ingresos que se encuentran por encima de la media. Este parámetro se actualiza anualmente de la siguiente manera:

$$\alpha_{t+1} = \frac{\widehat{Gini}_{t+1}}{Gini_t} - 1$$

Inicialmente las proyecciones de la CEPAL asumían que no había cambio distributivo ($\alpha = 0$) en períodos de aumento del PIB per cápita y que había deterioro distributivo en períodos de contracción. No obstante, en años recientes CEPAL ha asumido que $\alpha = 0$ tanto en períodos de crecimiento como de caída del PIB. A su vez, como se indicó anteriormente, el parámetro β surge de las proyecciones que hace CEPAL de la tasa de crecimiento del PIB per cápita (en dólares constantes de 2010). Sin embargo, este parámetro se va actualizando durante el año en la medida en que se actualizan las proyecciones de las tasas de crecimiento económico que realiza CEPAL. La proyección inicial de los ingresos para el año $t + 1$ se realiza usando las proyecciones de β publicadas por CEPAL en noviembre/diciembre de año t a través del reporte del "Balance Preliminar de las Economías de América Latina y el Caribe"⁵.

Resulta relevante resaltar la naturaleza "micro-macro" de esta estrategia, a fines de contextualizar la naturaleza de varios ejercicios que se implementarán en este trabajo. La parte "macro" del método se refiere al hecho de que la tasa de pobreza, el ingreso medio, y las variaciones proyectadas en el ingreso medio y en el coeficiente de Gini operan a nivel "macro", es decir, para un periodo y país. La parte "micro" de la metodología se refiere a que los pronósticos de pobreza surgen de agregar predicciones de los ingresos de cada persona, provenientes de encuestas de hogares. Más concretamente, la fórmula descrita más arriba parte de información "micro" (los ingresos de las personas), y genera una predicción (también "micro") de los ingresos de esas personas, sobre la base de la información "macro" (la requerida para computar α , β , y μ_t). Estas predicciones "micro" son luego agregadas para obtener una predicción de las tasas agregadas de pobreza y pobreza extrema, para un país y período determinado, de una naturaleza "macro" en la taxonomía propuesta en esta sección.

Consecuentemente, es importante señalar que la naturaleza del ejercicio predictivo implementado por la metodología de CEPAL tiene un fuerte contenido "macro", tanto en sus insumos, como las variaciones del ingreso medio y del Gini, como en su resultado, las tasas de pobreza agregada, y que la parte "micro" es esencialmente una regla de cálculo para, luego de actualizar los ingresos al nivel de los individuos, agregarlos para obtener una predicción macro. Esta apreciación resultará importante para explorar alternativas de mejora de este modelo base.

⁵ <https://www.cepal.org/es/publicaciones/bp>.

IV. Resultados y desempeño del modelo base en 2019

Este capítulo implementa la metodología base propuesta por CEPAL para los datos de las encuestas de hogares de 2018 (o la más reciente disponible hasta ese año), replicando las proyecciones realizadas para el año 2019 usando las últimas proyecciones de α y β . El Cuadro 4 reporta, para distintos países de la región, las proyecciones agregadas de las tasas de pobreza y pobreza extrema junto con el error de proyección expresado como diferencia en puntos porcentuales y como cambio porcentual respecto al dato observado, medido con la encuesta de hogares de 2019.

Cuadro 4
Proyecciones agregadas de pobreza y pobreza extrema para 2019^a
(En puntos porcentuales)

País	Pobreza			Pobreza extrema		
	Proy.	Obs.	Error	Proy.	Obs.	Error
Argentina	25,9	27,2	1,3 (4,8%)	3,8	4,2	0,4 (9,5%)
Bolivia (Estado Plurinacional de)	33,1	31,1	-2,0 (-6,4%)	14,3	12,1	-2,2 (-18,2%)
Brasil	19,4	19,2	0,2 (-1,0%)	5,4	5,5	0,1 (1,8%)
Chile	10,2	-	-	1,4	-	-
Colombia	29,5	-	-	10,8	-	-
Costa Rica	15,8	16,5	0,7 (4,2%)	3,9	3,4	-0,5 (-14,7%)
Ecuador	25,0	25,7	0,7 (2,7%)	6,8	7,6	0,8 (10,5%)
El Salvador	33,8	30,4	-3,4 (-11,2%)	7,5	5,6	-2,0 (-35,7%)
Guatemala	48,5	-	-	20,7	-	-
Honduras	55,1	52,3	-2,8 (-5,4%)	18,3	20,0	1,7 (8,5%)
México	42,2	-	-	11,2	-	-
Nicaragua	46,3	-	-	17,7	-	-

País	Pobreza			Pobreza extrema		
	Proy.	Obs.	Error	Proy.	Obs.	Error
Panamá	14,4	14,6	0,2 (1,4%)	6,7	6,6	-0,1 (-1,5%)
Paraguay	19,5	19,4	-0,1 (-0,5%)	6,3	6,2	-0,1 (-1,5%)
Perú	16,6	15,4	-1,2 (-7,8%)	3,7	3,0	-0,7 (-23,3%)
República Dominicana	20,7	20,3	-0,4 (-2,0%)	4,6	4,3	-0,3 (-7,0%)
Uruguay	2,9	3,0	0,1 (3,3%)	0,1	0,1	0,0 (0,0%)

Fuente: Elaboración propia.

^aEntre paréntesis se reportan los errores de pronóstico como cambio porcentual respecto del dato observado.

En términos generales, y teniendo en cuenta las limitaciones informativas bajo las cuales se realizan estas predicciones (que solo se basan en un mínimo de información disponible, de crecimiento y desigualdad), las predicciones presentan valores razonables, tanto a nivel absoluto como relativo. Es decir, para el caso de un país de muy elevada pobreza, como Honduras, el error de predicción es de tan solo 2.8 puntos, una cifra menor. En términos generales, los errores de pronósticos (medidos como diferencia absoluta) son mayores para la tasa de pobreza extrema que para la tasa de pobreza. Cuando comparamos entre países, los mayores errores se dan en El Salvador, Perú, Bolivia, Argentina y Costa Rica. Tanto para la pobreza extrema como para la pobreza, las proyecciones estuvieron por encima del valor observado en 7 casos, y por debajo en 5.

Para los 12 países para los cuales se pudo computar el error de pronóstico (al tener el dato observado de la encuesta de hogares para el año 2019) se calcularon las medidas tradicionales de desempeño medio de pronóstico como se reporta en el Cuadro 5, habituales en la literatura moderna de *machine learning*.

Las principales medidas de desempeño de pronóstico utilizadas son: el error medio (ME), la raíz del error cuadrático medio (RMSE), el error medio absoluto (MAE), el error medio porcentual (MPE), y el error medio absoluto porcentual (MAPE).

Cuadro 5
Principales medidas de desempeño de pronóstico

	Pobreza	Pobreza extrema
Error medio (ME)	-0,59	-0,24
Raíz del error cuadrático medio (RMSE)	1,52	1,06
Error medio absoluto (MAE)	1,09	0,74
Error medio porcentual (MPE)	-1,49	-5,97
Error medio absoluto porcentual (MAPE)	4,23	11,03

Fuente: Elaboración propia.

Estas medidas implican una evaluación agregada de las proyecciones realizadas para este conjunto de doce países. Dado que el error de pronóstico está definido como la diferencia entre el dato observado y la proyección de pobreza, errores medios negativos significan que se tiende a sobrepronosticar tanto la tasa de pobreza como la tasa de pobreza extrema. Las medidas expresadas en términos porcentuales (MPE y MAPE) muestran, nuevamente, un peor desempeño de los pronósticos de la tasa de pobreza extrema respecto a los pronósticos de pobreza. De todas formas, en términos económicos, son valores relativamente pequeños por lo que, en el promedio, hay un buen ajuste de las proyecciones realizadas por CEPAL para el año 2019 para el agregado de los países de la región.

Parte de la riqueza de la parte "micro" de la metodología es poder explotar los microdatos para proyectar las tasas de pobreza para ciertos grupos de interés, lo que permite realizar una evaluación a nivel desagregado de las proyecciones. Sin embargo, surge una dificultad propia de trabajar con los microdatos y

es que para el año t y $t + 1$ no se observan los mismos hogares, por lo que no se puede comparar si un hogar que en t es proyectado como pobre fue efectivamente pobre en $t + 1$. En principio para algunos países sería posible aprovechar la estructura de "panel rotativo" en sus esquemas muestrales para resolver la cuestión anterior. No obstante, la cantidad de observaciones disponibles para el panel se reduce dramáticamente, tanto porque los cambios solo son observables para la proporción de individuos que permanecen en la rotación, como por los problemas de abandono no sistemático del panel ("attrition"). Si a esto se agrega que dichas estructuras de panel rotativo no están disponibles para todos los países de la región, se concluye que el ejercicio basado en paneles es de una naturaleza muy restrictiva.

Ante la imposibilidad efectiva de realizar una evaluación desagregada a nivel de las personas u hogares, trabajamos con un grado mayor de agregación construyendo "pseudo-individuos", segmentando a las personas en 90 grupos (potenciales) en función de: su condición de actividad (Ocupado, Desocupado, Inactivo, NA⁶, NS/NR), género (Mujer, Hombre), nivel educativo (Alto = más de 12 años de educación formal, Medio = entre 8 y 12 años, Bajo = menos de 8 años), y grupo etario (menores de 18 años, entre 18 y 64 años, de 65 años o más). Entonces, estos potenciales 90 "pseudo-individuos" surgen de todas las posibles combinaciones entre estas cuatro variables de interés. Para cada grupo se calcula la tasa de pobreza proyectada y observada (no así la tasa de pobreza extrema)⁷ y se determina si cada "pseudo-individuo" se encuentra en condición de pobreza o no tanto en función del ingreso medio como en función del ingreso mediano.

Esta construcción de "pseudo-individuos" y la evaluación de los respectivos pronósticos se realizó para cada uno de los 12 países. Para cada país se calcula la "matriz de confusión" y dos medidas de desempeño que resultan de particular interés en este caso: sensibilidad o exhaustividad (*recall*) y precisión (*precision*). La medida de sensibilidad también es conocida como la tasa de verdaderos positivos y se calcula como el cociente de verdaderos positivos sobre el total de positivos. Esta medida responderá a la pregunta "del total de casos observados en situación de pobreza, ¿qué porcentaje se predice correctamente como tal?". Por otro lado, la medida de precisión se calcula como el porcentaje de casos positivos detectados y nos permitirá responder a la pregunta "de los predichos pobres, ¿qué porcentaje es realmente pobre?".

Junto con estas dos métricas se reporta el F1-Score que es la media armónica entre sensibilidad y precisión. Esta medida resulta especialmente útil cuando la distribución entre pobres y no-pobres es desigual, como en estos casos.

En el cuadro 6 se reportan las medidas de sensibilidad, precisión y F1-Score para cada uno de los 12 países, para la tasa de pobreza calculada usando el ingreso medio de los pseudo-individuos, mientras que en anexo 1 se muestran las matrices de confusión de cada país.

Cuadro 6
Evaluación desagregada de las proyecciones de 2019

País	Agregación en media			Agregación en mediana		
	Sensibilidad	Precisión	F ₁ -Score	Sensibilidad	Precisión	F ₁ -Score
Argentina	66,7%	66,7%	0,667	60,7%	65,4%	0,630
Bolivia (Estado Plurinacional de)	83,3%	62,5%	0,714	94,1%	76,2%	0,842
Brasil	66,7%	66,7%	0,667	85,7%	85,7%	0,857
Costa Rica	0,00%	0,00%	0,000	25,0%	66,7%	0,363
Ecuador	25,0%	100,0%	0,400	80,0%	88,9%	0,842
El Salvador	83,3%	55,6%	0,667	92,9%	68,4%	0,788

⁶ En el caso de la condición de actividad, la categoría NA corresponde a niños y jóvenes que, por su edad, no integran la población en edad de trabajar.

⁷ Esto se debe a que al haber tasas de pobreza extrema relativamente bajas para varios países y estar agregando los microdatos en 90 grupos, en muchos casos no se identifican individuos cuyos ingresos (medios o medianos) estén por debajo de la línea de indigencia.

País	Agregación en media			Agregación en mediana		
	<i>Sensibilidad</i>	<i>Precisión</i>	<i>F₁-Score</i>	<i>Sensibilidad</i>	<i>Precisión</i>	<i>F₁-Score</i>
Honduras	73,1%	86,4%	0,792	87,8%	87,8%	0,878
Panamá	33,3%	33,3%	0,333	60,0%	60,0%	0,600
Paraguay	33,3%	50,0%	0,400	50,0%	60,0%	0,545
Perú	0,00%	0,00%	0,000	50,0%	27,3%	0,353
República Dominicana ^a	--	--	--	50,0%	40,0%	0,444
Uruguay ^a	--	--	--	--	--	--

Fuente: Elaboración propia.

^aPara República Dominicana (agregación en media) y Uruguay no se identifican grupos pobres.

Una combinación de alta sensibilidad y alta precisión (como en Honduras) implica que el modelo maneja muy bien esta clasificación de la situación de pobreza ya que logra un alto ratio de positivos detectados (personas pobres) y un alto porcentaje de clasificaciones correctas de personas verdaderamente pobres. Una baja sensibilidad pero con alta precisión (como en Ecuador agregando en función del ingreso medio o Costa Rica, agregando en función del ingreso mediano) implica que el modelo no detecta muy bien la clasificación de la situación de pobreza, pero cuando lo hace es altamente confiable. Una alta sensibilidad pero con baja precisión (que no se da en estos países) implicaría que el modelo detecta bien la clasificación de pobres, pero también incluye muestras de la clasificación de no-pobres. Por último, un valor bajo en ambos indicadores (como en Perú o Costa Rica) indica que el modelo no logra clasificar a los pobres correctamente ya que tiene un bajo ratio de positivos detectado y un bajo porcentaje de clasificaciones correctas de personas realmente pobres.

En términos generales, para todos los países, a excepción de Argentina, las distintas medidas reportadas mejoran cuando la agregación se hace calculando la mediana del ingreso. Los países en los que se identifican correctamente más grupos pobres son Honduras, Brasil, Bolivia y Ecuador, mientras que las proyecciones no tienen buen desempeño en Perú, Costa Rica y Paraguay.

V. Extensiones del modelo base y resultados para 2019

En una primera instancia, se buscó una mínima flexibilización del modelo base propuesto por CEPAL de forma tal de trabajar con modelos dinámicos en panel que arrojen proyecciones de α y β usando la información mínima requerida por el modelo base. Este capítulo explota explícitamente la naturaleza “micro-macro” de la metodología usada por CEPAL. Desde una perspectiva “macro”, para cada país se observa la serie temporal de ingresos medios, coeficientes de Gini y tasas de pobreza agregadas, aun cuando esto sea imposible a nivel micro, teniendo en cuenta que, como adelantásemos en el capítulo anterior, los ingresos provienen de encuestas de hogares que, salvo excepciones y por periodos muy cortos, tienen una estructura de corte transversal y no de panel. Entonces, esta sección explota la variación geográfica y temporal (de panel) del componente “macro” para refinar las predicciones, a fines de explotar la estructura dinámica de la información a ese nivel.

Dado nuestro panel de 12 países a lo largo de 20 años (2000-2019), buscamos desarrollar modelos que permitan predecir tanto la tasa de crecimiento del PIB per cápita (medida como diferencia logarítmica del PIB per cápita) como del nivel del coeficiente de Gini. Para evaluar que la dinámica es efectivamente relevante en la determinación de ambas variables, se estimó un modelo simple del tipo autorregresivo de orden uno (AR(1)) en panel usando el estimador de Arellano-Bond para cada variable. La especificación del modelo obedece a razones puramente pragmáticas: con solo 19 años, incorporar más rezagos implica perder 12 observaciones por cada uno de ellos. A continuación, se presentan los modelos estimados (entre paréntesis se reportan los errores estándares):

$$\Delta \ln(\widehat{pib}_{it}) = \frac{0.02}{(0.002)} + \frac{0.27}{(0.07)} \Delta \ln(pib_{it-1})$$

$$\widehat{Gini}_{it} = \frac{0.35}{(0.03)} - \frac{0.003}{(0.0003)} t + \frac{0.35}{(0.06)} Gini_{it-1}$$

Cabe señalar que el modelo del coeficiente de Gini incluye una tendencia lineal determinística para controlar por el comportamiento de persistencia en el tiempo. Incluso controlando por dicha tendencia el término autorregresivo es estadísticamente significativo a valores tradicionales de significación en ambas estimaciones. En otras palabras, la dinámica es relevante en la determinación de dichas variables por lo que se justifica el uso de modelo dinámicos para la estimación y proyección de estas variables.

Entonces, restringimos el panel hasta 2018 y realizamos proyecciones de ambas variables para 2019 a partir de los siguientes modelos:

- Modelo 1: se estima un vector autorregresivo para paneles (PVAR), PVAR(1) entre $\Delta \ln(pib_{it})$ y $Gini_{it}$ por método de los momentos generalizado (GMM por sus siglas en inglés), considerando que el panel es desbalanceado y usando errores estándares clusterizados por país⁸.
- Modelo 2: Modelo 1 que incluye un término cuadrático del $\Delta \ln(pib_{it-1})$ en forma exógena.
- Modelo 3: Modelo 2, pero se seleccionan los rezagos, término cuadrático y efectos fijos por país a través de una regularización del tipo LASSO⁹, buscando el factor de penalización, λ , que minimice el criterio de información de AIC. LASSO selecciona los siguientes regresores para cada una de las variables modeladas:
 - Para $\Delta \ln(pib_{it})$: su término autorregresivo, el primer rezago del coeficiente de Gini, el primer rezago del término cuadrático del ingreso, y efectos fijos para Brasil, Panamá, Perú y República Dominicana.
 - Para $Gini_{it}$: sólo el término autorregresivo resulta seleccionado.

Todos los modelos, al incluir los rezagos, buscan captar la dinámica temporal propia de estas variables. El modelo 1, el PVAR, no solo considera la dinámica propia de cada variable (tasa de crecimiento del PIB y del Gini) sino también la dinámica de la otra. El modelo 2 busca capturar alguna no linealidad en el ingreso a partir de la incorporación, en forma exógena, de un término cuadrático de la tasa de crecimiento del producto rezagada un período. El modelo 3 utiliza técnicas de regularización, como LASSO, para elegir modelos más parsimoniosos partiendo de un modelo más general, lo cual es relevante en este caso, considerando que solo se dispone de 20 observaciones temporales.

Como prueba adicional, se evaluó la selección de las variables a partir del método de regresión *Elastic Net*, pero los resultados obtenidos fueron los mismos que por LASSO, lo cual es esperable, teniendo en cuenta que la cantidad de regresores es baja y relativamente ortogonal entre sí, de modo que LASSO no enfrenta los problemas de “agrupamiento” que resuelve *Elastic Net*.

El cuadro 7 reporta los pronósticos de ambas variables para 2019 de cada uno de los tres modelos para los 12 países.

En el cuadro 7 se comparan los pronósticos de los tres modelos y se destaca en negrita aquellos que se acercan más al valor observado. Podemos observar una clara ganancia en los pronósticos al cuando se utiliza LASSO para seleccionar los regresores a incluir. Esto reporta una ventaja implica estimar un modelo más parsimonioso dados los grados de libertad disponibles.

⁸ Se utiliza el comando `pvar` de STATA.

⁹ Ver Elliott, G. y Timmerman, A., 2016, *Economic Forecasting*, Princeton University Press, New York, para una descripción detallada de LASSO, elastic net y otras estrategias de regularización, y de los criterios de evaluación como AIC (*Akaike Information Criterion*), en particular para su uso en el contexto de paneles y series temporales.

Cuadro 7
Pronósticos para 2019 de crecimiento ($\Delta \ln(pib_{it})$) y desigualdad ($Gini_{it}$)^a

País	Tasa de crecimiento del PIB pc				Coeficiente de Gini			
	Observado	Pronóstico			Observado	Pronóstico		
		PVAR	PVAR+ exog	LASSO		PVAR +	PVAR + exog	LASSO
		(1)	(2)	(3)		(1)	(2)	(3)
Argentina	-0,031	0,037	0,029	0,001	0,400	0,387	0,388	0,406
Bolivia (Estado Plurinacional de)	0,008	0,063	0,045	0,018	0,430	0,426	0,430	0,442
Brasil	0,006	0,070	0,047	0,010	0,538	0,526	0,530	0,530
Colombia	0,019	0,067	0,045	0,017	0,529	0,507	0,511	0,512
Costa Rica	0,011	0,066	0,045	0,018	0,495	0,480	0,484	0,489
Ecuador	-0,016	0,054	0,036	0,010	0,456	0,443	0,446	0,456
El Salvador	0,018	0,056	0,039	0,013	0,406	0,394	0,398	0,413
Panamá	0,013	0,067	0,046	0,027	0,492	0,479	0,483	0,488
Paraguay	-0,013	0,065	0,045	0,018	0,473	0,462	0,465	0,473
Perú	0,005	0,061	0,043	0,026	0,429	0,428	0,431	0,443
República Dominicana	0,039	0,073	0,061	0,043	0,432	0,429	0,431	0,444
Uruguay	-0,001	0,052	0,036	0,010	0,392	0,381	0,384	0,401

Fuente: Elaboración propia.

^aSe resaltan los pronósticos que más se acercan al valor observado.

A partir de estos pronósticos de $\Delta \ln(pib_{it})$ y $Gini_{it}$ para 2019 se utiliza el modelo base desarrollado por CEPAL para pronosticar la pobreza usando microdatos a partir de la proyección del ingreso tal que:

$$\widehat{y}_i^{2019} = (1 + \hat{\beta}_{2019})[(1 - \hat{\alpha}_{2019})y_i^{2018} + \hat{\alpha}_{2019}\mu^{2018}]$$

donde $\hat{\beta}_{2019}$ es la proyección de $\Delta \ln(pib_{it}) \times 100$ de cada modelo y $\hat{\alpha}_{2019} = \frac{\widehat{Gini}_{2019}}{Gini_{2018}} - 1$.

El cuadro 8 reporta los $\hat{\beta}_{2019}$ y $\hat{\alpha}_{2019}$ proyectados de cada modelo y los compara, a su vez, con los que CEPAL realizó a fines de 2018 a partir de la publicación del "Balance preliminar de las economías de ALC".

Nuevamente, el modelo 3 seleccionado por LASSO pareciera ser una buena alternativa para pronosticar la tasa de crecimiento anual del PIB per cápita, incluso cuando se lo compara con las proyecciones para 2019 hechas por CEPAL a fines de 2018. Por otro lado, si bien CEPAL en los últimos años asume que $\alpha = 0$ para todos los países, no está claro que esto necesariamente sea lo mejor para cada uno de los países.

Se utilizan, entonces, los valores pronosticados de $\hat{\beta}_{2019}$ y $\hat{\alpha}_{2019}$ por cada modelo para proyectar los ingresos de 2019 a partir de los datos de la encuesta de hogares de 2018 y, en función de las respectivas líneas de pobreza e indigencia, calcular la tasa de pobreza y tasa de pobreza extrema.

Cuadro 8
Comparación de los valores pronosticados de crecimiento ($\hat{\beta}_{2019}$) y cambio en la desigualdad ($\hat{\alpha}_{2019}$)^a

País	$\hat{\beta}_{2019}$					$\hat{\alpha}_{2019}$				
	Actual	Forecast				Actual	Forecast			
		CEPAL	PVAR	PVAR+ exog	LASSO		CEPAL	PVAR	PVAR+ exog	LASSO
	(0)	(1)	(2)	(3)		(0)	(1)	(2)	(3)	
Argentina	-3,051	-1,800	3,687	2,896	0,127	0,010	0,000	-0,024	-0,020	0,024
Bolivia (Estado Plurinacional de)	0,793	4,300	6,286	4,493	1,847	-0,018	0,000	-0,026	-0,019	0,009
Brasil	0,650	2,000	6,965	4,713	1,023	-0,004	0,000	-0,026	-0,018	-0,019
Colombia	1,851	3,300	6,711	4,541	1,744	0,017	0,000	-0,026	-0,018	-0,014
Costa Rica	1,104	2,900	6,602	4,533	1,787	0,004	0,000	-0,026	-0,018	-0,008
Ecuador	-1,626	0,900	5,420	3,641	0,992	0,004	0,000	-0,025	-0,018	0,003
El Salvador	1,847	2,400	5,607	3,899	1,336	0,002	0,000	-0,026	-0,018	0,020
Panamá	1,310	-2,000	6,686	4,629	2,721	0,029	0,000	-0,026	-0,018	-0,007
Paraguay	-1,295	4,200	6,495	4,513	1,802	-0,002	0,000	-0,026	-0,018	-0,003
Perú	0,511	3,600	6,121	4,288	2,649	-0,023	0,000	-0,026	-0,019	0,008
República Dominicana	3,882	5,700	7,304	6,067	4,254	-0,020	0,000	-0,027	-0,022	0,008
Uruguay	-0,138	1,500	5,218	3,563	1,038	0,003	0,000	-0,026	-0,018	0,026

Fuente: Elaboración propia.

^aPara CEPAL, $\alpha=0$ en los últimos dos años para evitar sobreestimar la pobreza.

El cuadro 9 reporta los pronósticos de las tasas de pobreza y pobreza extrema para 2019 para cada uno de los 12 países analizados usando los tres modelos de pronóstico propuestos e incluyendo los pronósticos agregados (modelo 4) y comparándolos con el realizado por CEPAL.

Se puede observar que en la estimación de la tasa de pobreza, el modelo 3 que selecciona por LASSO domina en 8 de los 12 países, mientras que en el caso de la tasa de pobreza extrema, el modelo 1 que estima un PVAR(1) por GMM domina en 7 de 12 países, seguido por el modelo 3 (selección por LASSO) en 6 de 12.

En resumen, aunque en esta instancia no se puede evaluar si dichas diferencias son estadísticamente significativas ya que estamos pronosticando un único año, se puede apreciar que habría ganancias en generar modelos dinámicos de pronóstico de los parámetros β y α del modelo base de CEPAL.

Por otro lado, los pronósticos anteriores fueron comparados con los pronósticos que surgen de modelar directamente las tasas de pobreza en función de la tasa de crecimiento del PIB per cápita y del coeficiente de Gini, es decir, sin usar la forma funcional que establece el modelo base. Llamaremos, entonces, modelo 4 a este modelo que genera pronósticos agregados de pobreza (sin usar los microdatos) que se estima a través del PVAR(1), pero seleccionado los rezagos y efectos fijos por país a través de LASSO. Para la selección de las variables del modelo, el valor del parámetro de penalización, λ , fue seleccionado de forma tal que minimice el criterio de información de Akaike (AIC).

Para la tasa de pobreza, las variables seleccionadas fueron: su término autorregresivo, el primer rezago de la tasa de crecimiento del PIB per cápita y del coeficiente de Gini y 10 efectos fijos (todos los países, menos Argentina y República Dominicana). Esos dos países no seleccionados pueden pensarse como casos atípicos.

Para la tasa de pobreza extrema, las variables seleccionadas fueron: su término autorregresivo, el primer rezago de la tasa de crecimiento del PIB per cápita y del coeficiente de Gini y 10 efectos fijos (todos los países, menos Argentina y Colombia). Nuevamente, estos dos países pueden interpretarse como casos atípicos.

No se observan claras ganancias por pronosticar las tasas de pobreza en forma agregada respecto de cualquier otra especificación que trabaje con los microdatos, es decir, respecto de otros enfoques que requieren proyectar primero la tasa de crecimiento del ingreso y el coeficiente de Gini. No hay diferencias significativas si se proyectan directamente las tasas de pobreza.

A continuación, presentamos algunas pruebas adicionales que realizamos sobre los pronósticos de la tasa de crecimiento del PIB per cápita y coeficiente de Gini realizado a través de LASSO para utilizar en las micro simulaciones del modelo base de CEPAL.

A. Efectos asimétricos

Los resultados de los capítulos anteriores sugieren que los modelos estándar tienen alguna dificultad en predecir en el caso de países con crecimiento negativo del PIB, como Argentina. A fines de focalizar en esta cuestión, se incorporaron en el modelo de tasa de crecimiento del PIB per cápita (en panel) dos términos asimétricos:

- i) Una variable dicotómica que toma el valor 1 si la tasa de crecimiento del PIB per cápita del año anterior (para cada país) resultó negativa (contracción económica), y toma el valor 0 en caso contrario.
- ii) Una variable dicotómica que toma el valor 1 si la tasa de crecimiento del PIB per cápita del año anterior (para cada país) estuvo por debajo de la media de los países de la región, y toma el valor 0 en caso contrario.

Cuadro 9
Comparación de pronósticos de las tasas de pobreza para 2019^a
(En porcentajes)

País	Tasa de pobreza						Tasa de pobreza extrema					
	Obs.	Pronóstico					Obs.	Pronóstico				
		CEPAL	PVAR	PVAR+ exog	LASSO	LASSO agregado		CEPAL	PVAR	PVAR+ exog	LASSO	LASSO agregado
	(0)	(1)	(2)	(3)	(4)	(0)	(1)	(2)	(3)	(4)		
Argentina	27,2	25,5	23,0	23,3	24,4	19,3	4,2	3,7	3,5	3,5	3,6	3,2
Bolivia (Estado Plurinacional de)	31,1	31,7	30,8	31,6	32,7	33,2	12,1	14,1	13,8	14,1	14,5	15,8
Brasil	19,2	19,0	17,6	18,3	19,3	20,4	5,5	5,3	4,9	5,1	5,4	5,2
Colombia	31,7	29,1	29,6	28,6	29,6	30,0	12,8	10,6	10,8	10,3	10,8	12,1
Costa Rica	16,5	15,5	14,4	14,8	15,7	16,0	3,4	3,9	3,8	3,8	3,9	3,7
Ecuador	25,7	23,7	21,9	22,4	23,7	22,5	7,6	6,4	5,8	6,0	6,4	6,4
El Salvador	30,4	33,5	32,0	32,8	34,1	35,1	5,6	7,5	6,9	7,2	7,7	7,1
Panamá	14,6	15,1	13,3	13,6	14,1	15,6	6,6	7,0	6,4	6,5	6,7	7,4
Paraguay	19,4	17,7	17,2	17,6	18,4	21,4	6,2	5,7	5,3	5,5	6,0	6,7
Perú	15,4	15,7	14,9	15,4	16,0	19,2	3,0	3,7	3,2	3,3	3,5	5,6
República Dominicana	20,3	19,9	19,3	19,3	20,6	25,6	4,3	4,4	4,2	4,2	4,6	6,7
Uruguay	3,0	2,7	2,3	2,3	2,8	2,8	0,1	0,1	0,1	0,1	0,1	-0,4

Fuente: Elaboración propia.

^aSe resaltan los pronósticos que más se acercan al valor observado.

Ambos términos de asimetría fueron incorporados en el modelo en panel de la tasa de crecimiento económica y se aplicó LASSO para la selección del modelo de mejor ajuste. Ninguno de los términos asimétricos resultó seleccionado. El resultado es razonable, toda vez que muchos de los modelos implementados explícita o implícitamente utilizan efectos fijos (porque se estiman a nivel de país o porque los incorporan en un agregado), de modo que la naturaleza "asimétrica" de la evolución de variables como el PIB es muy colineal al efecto fijo. Por ejemplo, son muy pocos países para los cuales se "activa" la variable binaria que capta el crecimiento negativo, confundándose con el efecto fijo por país. El efecto fijo captura cuestiones idiosincráticas de los países que inciden en forma diferencial respecto del resto de la región en la determinación del crecimiento o del Gini. Es importante remarcar que en este caso el rol de los efectos fijos es ver si los mismos ayudan a mejorar las predicciones condicionales por país, a diferencia del uso estándar de los mismos en econometría tradicional, donde su incorporación se debe a evitar sesgos por omisión de variables que no varían por países.

B. Incorporación de información externa

Una segunda prueba consistió en incorporar al PVAR(1) algunas variables macroeconómicas externas que puedan impactar sobre las variables predichas, tales como:

- i) La tasa de crecimiento anual de un índice de precios reales de commodities o productos básicos vinculadas a la energía obtenida del Banco Mundial (The Pink Sheet). Este índice incluye materias primas tales como el carbón, el petróleo crudo, el gas natural y el gas natural licuado.
- ii) La tasa de crecimiento anual de un índice de precios reales de commodities agrícolas obtenida del Banco Mundial (The Pink Sheet). Este índice incluye materias primas tales como cacao, café, té, aceites (de coco, palma, soja), poroto de soja, harina de soja, cebada, maíz, arroz, trigo, banana, carne (vacuna y bovina), naranjas, camarones, azúcar, algodón, tabaco, madera y caucho.
- iii) La tasa de crecimiento anual del tipo de cambio nominal de cada uno de los países obtenida de CEPALSTAT y complementada con estadísticas oficiales (en el caso de Argentina usando datos del Banco Central de la República Argentina).

Ambos índices de precios reales de materias primas se comportan como efectos fijos temporales (cambian a lo largo del tiempo, pero son constantes por país), mientras que la variación del tipo de cambio nominal tiene heterogeneidad temporal y por país.

Las tres variables fueron incorporadas en el PVAR(1) y se hizo una selección por LASSO tanto para la modelación de la tasa de crecimiento del producto como del nivel del coeficiente de Gini. Las tres variables resultaron seleccionadas únicamente para los modelos de tasa de crecimiento del PIB per cápita, no así para el Gini, lo cual es razonable, teniendo en cuenta que el índice de Gini tiene un comportamiento tendencial relativamente suave, por ende, menos sensible a variaciones cíclicas de corto plazo como las que se pretende captar con estas variables. El cuadro 10 muestra la comparación de este nuevo modelo (nuevo Modelo 5).

Solo se observa una ganancia en los pronósticos de la tasa de crecimiento del producto para Argentina y El Salvador. No resulta dominante respecto del Modelo 3 seleccionado por LASSO que no incluye estos regresores exógenos.

Cuadro 10
Comparación de los pronósticos de $\hat{\beta}_{2019}$ y $\hat{\alpha}_{2019}$ incorporando variables externas^a

País	$\hat{\beta}_{2019}$						$\hat{\alpha}_{2019}$				
	Actual	Forecast					Actual	Forecast			
		CEPAL	PVAR	PVAR+ exog	LASSO	LASSO + i. externa		CEPAL	PVAR	PVAR+ exog	LASSO
	(0)	(1)	(2)	(3)	(5)		(0)	(1)	(2)	(3)	
Argentina	-3,051	-1,800	3,687	2,896	0,127	-3,351	0,010	0,000	-0,024	-0,020	0,024
Bolivia (Estado Plurinacional de)	0,793	4,300	6,286	4,493	1,847	3,136	-0,018	0,000	-0,026	-0,019	0,009
Brasil	0,650	2,000	6,965	4,713	1,023	-0,257	-0,004	0,000	-0,026	-0,018	-0,019
Colombia	1,851	3,300	6,711	4,541	1,744	2,377	0,017	0,000	-0,026	-0,018	-0,014
Costa Rica	1,104	2,900	6,602	4,533	1,787	2,444	0,004	0,000	-0,026	-0,018	-0,008
Ecuador	-1,626	0,900	5,420	3,641	0,992	1,728	0,004	0,000	-0,025	-0,018	0,003
El Salvador	1,847	2,400	5,607	3,899	1,336	1,761	0,002	0,000	-0,026	-0,018	0,020
Panamá	1,310	-2,000	6,686	4,629	2,721	3,676	0,029	0,000	-0,026	-0,018	-0,007
Paraguay	-1,295	4,200	6,495	4,513	1,802	2,584	-0,002	0,000	-0,026	-0,018	-0,003
Perú	0,511	3,600	6,121	4,288	2,649	3,390	-0,023	0,000	-0,026	-0,019	0,008
República Dominicana	3,882	5,700	7,304	6,067	4,254	4,497	-0,020	0,000	-0,027	-0,022	0,008
Uruguay	-0,138	1,500	5,218	3,563	1,038	2,302	0,003	0,000	-0,026	-0,018	0,026

Fuente: Elaboración propia.

^aPara CEPAL, $\alpha=0$ en los últimos dos años para evitar sobreestimar la pobreza.

C. Pronósticos de pobreza para los países faltantes

Tal como se mencionó anteriormente, en la construcción de la base de datos en panel se eliminaron seis países para los cuales no había suficientes datos temporales de PIB per cápita, coeficiente de Gini y tasas de pobreza entre 2000 y 2019. Estos son: Chile, Guatemala, Honduras, México, Nicaragua y Venezuela. A la fecha, la última encuesta de hogares disponible para Chile es de 2017, para Guatemala, Nicaragua y Venezuela es de 2014, mientras que para Honduras y México es de 2018.

El cuadro 11 reporta los pronósticos de β (la tasa de crecimiento del producto real) y α (variación porcentual del Gini) de cada uno de los modelos utilizados. Notar que no se observa en 2019 la variación respecto de 2018 del Gini de muchos países. Para los modelos (1), (2) y (3) se utilizó la última información disponible en CEPALSTAT del Gini de cada país.

Cuadro 11
Pronósticos para 2019 de β y α para los países faltantes^a

País	$\hat{\beta}_{2019}$					$\hat{\alpha}_{2019}$				
	Actual	Pronóstico				Actual	Pronóstico			
		CEPAL	PVAR	PVAR+ exog	LASSO		CEPAL	PVAR	PVAR+ exog	LASSO
	(0)	(1)	(2)	(3)	(0)	(1)	(2)	(3)		
Chile	-0,134	3,300	6,116	4,339	1,713	NA	0,000	-0,011	-0,008	0,011
Guatemala	1,875	3,000	6,582	4,474	1,711	NA	0,000	-0,013	0,028	0,068
Honduras	0,982	3,600	6,577	4,566	1,840	0,013	0,000	-0,012	-0,009	-0,004
México	-1,139	2,100	6,171	4,182	1,480	NA	0,000	-0,012	-0,009	-0,003
Nicaragua	-5,054	-2,000	3,711	1,468	0,751	NA	0,000	-0,010	-0,006	0,006

Fuente: Elaboración propia.

^aNA indica que no se dispone de la variación anual entre 2019 y 2018 del Gini.

En el cuadro 12 se reportan las tasas de pobreza pronosticadas para 2019 a partir de los distintos modelos estimados usando la última encuesta de hogares disponible y los valores de α y β proyectados. A excepción de Chile, las tasas de pobreza pronosticadas para 2019 para los países faltantes están por encima del 40%.

Cuadro 12
Pronósticos de tasas de pobreza de 2019 para los países faltantes^a
(En porcentajes)

País	Tasa de pobreza					Tasa de pobreza extrema				
	Actual	Pronóstico				Actual	Pronóstico			
		CEPAL	PVAR	PVAR+ exog	LASSO		CEPAL	PVAR	PVAR+ exog	LASSO
	(0)	(1)	(2)	(3)	(0)	(1)	(2)	(3)		
Chile	NA	10,2	9,2	9,6	10,4	NA	1,4	1,3	1,3	1,4
Guatemala	NA	48,5	48,6	49,7	50,8	NA	20,7	20,8	21,4	22,2
Honduras	52,3	53,7	52,0	53,1	54,6	20,0	17,3	16,2	16,9	18,1
México	NA	40,3	38,2	39,3	40,7	NA	10,4	9,6	10,0	10,5
Nicaragua	NA	46,3	44,5	45,3	45,6	NA	17,7	16,4	16,8	17,1

Fuente: Elaboración propia.

^aNA indica que no se dispone del dato observado para 2019.

VI. Evaluación del desempeño predictivo para el período 2003 a 2019

Para validar los resultados obtenidos a partir de los modelos de pronóstico desarrollados sobre los datos de pobreza de 2019, llevamos a cabo un ejercicio de validación cruzada. Dicho ejercicio consistió en quitar del panel de datos un año a la vez, entre 2003 y 2019¹⁰, pronosticar las tasas de pobreza correspondientes de cada país y luego evaluar el desempeño de los pronósticos de los distintos modelos para dicho período temporal. En otras palabras, se trabaja con el panel original de datos y, en forma iterativa, se extrae un año a la vez que es usado como conjunto de validación de los modelos.

Llamaremos,

- **Modelo 0:** a las proyecciones de CEPAL para el año t usando proyecciones del PIB per cápita realizadas en $t - 1$.
- **Modelo 1:** al PVAR(1) entre la tasa de crecimiento del PIB per cápita y el coeficiente de Gini.
- **Modelo 2:** al PVAR(1) que a su vez incluye un término cuadrático del primer rezago de la tasa de crecimiento del producto, en forma exógena.
- **Modelo 3:** al PVAR(1) cuyos regresores y efectos fijos son seleccionados a través de LASSO.
- **Modelo 4:** pronósticos agregados de pobreza a partir de un PVAR(1) usando LASSO para la selección de rezagos y efectos fijos.
- **Modelo 5:** ídem Modelo 3, pero que incluye como regresores exógenos información externa tal como la tasa de crecimiento del precio real de los productos básicos agrícolas, de energía y del tipo de cambio nominal.

¹⁰ No se replicó el ejercicio para los primeros años del panel porque en esos años hay una mayor proporción de datos perdidos en la muestra.

Junto con estos modelos se evaluó una combinación (*pooling*) de pronósticos entre el modelo 0 y el 3, y entre el modelo 0 y el 5, tomando promedios simples. El principal objetivo de realizar una combinación de pronósticos es evaluar si se logra una mayor reducción del sesgo y la varianza obteniendo un modelo híbrido que resulte dominante respecto de los modelos individuales. La combinación de pronósticos resulta especialmente relevante cuando los modelos de pronósticos tienen problemas de especificación o el proceso generador de los datos cambia en el tiempo (Hendry y Clements, 2001), lo cual resulta relevante en este caso.

Los cuadros 13 a 16 reportan distintas medidas de desempeño de pronóstico de los modelos enunciados anteriormente para los 12 países del panel entre 2003 y 2019 (17 observaciones temporales). Las medidas de desempeño que se reportan son el error medio de pronóstico, la raíz del error cuadrático medio (RMSE), el error medio absoluto (MAE) y el error mediano absoluto (MdAE).

A su vez, se evaluó si las diferencias observadas en estas medidas de desempeño resultaron estadísticamente significativas entre modelos, siempre se evaluó el pronóstico de CEPAL con respecto al mejor o segundo mejor modelo.

En el cuadro 13, el error medio de pronóstico se reporta como una forma de evaluar el sesgo que cometen cada uno de los modelos. Las diferencias fueron evaluadas a partir de los valores absolutos, es decir, tratando en forma simétrico a los sesgos negativos y positivos.

En términos generales, se observa que, para los pronósticos de la tasa de pobreza, las proyecciones de CEPAL no siempre reportan un sesgo estadísticamente inferior que el resto de los modelos. CEPAL logra menores sesgos para Bolivia, Brasil y Uruguay y dichos sesgos son negativos lo cual implica que tiende a sobrepronosticar levemente la tasa de pobreza de esos países. El modelo (3) estimado por LASSO o una combinación entre las proyecciones de CEPAL y LASSO pueden reportar ganancias para el resto de los países. Los resultados son más dispares cuando se analizan los sesgos de los pronósticos de la tasa de pobreza extrema.

En el cuadro 14, prácticamente no se encuentran diferencias estadísticamente significativas entre las proyecciones de CEPAL y del resto de los modelos de pronósticos. Una excepción es Argentina. Para este país se encuentra evidencia de una mejora en el desempeño de los pronósticos de la tasa de pobreza cuando se utiliza LASSO para la selección de predictores en los modelos en los cuales también se incluyen variables macroeconómicas exógenas tales como las tasas de crecimiento de los precios de las materias primas y del tipo de cambio nominal.

De todas formas, podemos observar que trabajar con las proyecciones de CEPAL o alguna combinación de dichas proyecciones con las arrojadas por LASSO reportaría algún tipo de ganancia predictiva en 9 de los 12 países, para el caso de la tasa de pobreza, y en 6 de los 12 países, para el caso de la tasa de pobreza extrema.

Cuando se trabaja con una función de pérdida absoluta (como el error medio absoluto) en el cuadro 15, tampoco se encuentra evidencia estadística de diferencias significativas entre las proyecciones de CEPAL y el mejor (o segundo mejor) modelo alternativo. La excepción nuevamente es Argentina donde se encuentra una clara ganancia por incorporar variables macroeconómicas exógenas en la estimación de los modelos.

Cuadro 13
Error medio de pronóstico de los distintos modelos (2003-2019)^a

País	Error medio															
	Pobreza								Pobreza extrema							
	CEPAL	PVAR	PVAR+ exog	LASSO	LASSO agregado	LASSO+ exog	Pooling	Pooling	CEPAL	PVAR	PVAR+ exog	LASSO	LASSO agregado	LASSO+ exog	Pooling	Pooling
(0)	(1)	(2)	(3)	(4)	(5)	(0)+(3)	(0)+(5)	(0)	(1)	(2)	(3)	(4)	(5)	(0)+(3)	(0)+(5)	
Argentina	0,087	1,760	1,467	-0,833	-2,763	-0,413	-0,373	-0,163	-0,973	-0,813	-0,827	-1,133	-1,377	-1,033	-1,053	-1,003
Bolivia (Estado Plurinacional de)	-0,372 ^d	1,655	0,948	-1,158	-0,775	-0,965	-0,765	-0,668	-0,788	0,472	0,052 ^d	-1,241	-0,867	-1,128	-1,015	-0,958
Brasil	-0,249 ^b	2,070	1,433	-0,586	-1,055	-0,655	-0,417	-0,452	0,074	0,605	0,461	0,005 ^d	-0,006	-0,033	0,039	0,020
Colombia	0,090	2,530	1,790	-0,423	-1,521	-0,383	-0,167	-0,147	-0,033 ^c	1,260	0,840	-0,407	-1,244	-0,327	-0,220	-0,180
Costa Rica	0,606	2,282	1,788	0,141 ^d	-0,914	0,200	0,374	0,403	0,065	0,441	0,312	-0,047	-0,236	-0,029	0,009^d	0,018
Ecuador	0,253	2,713	1,926	-0,467	-1,462	-0,394	-0,107	-0,071 ^c	0,049	1,036	0,703	-0,224	-0,627	-0,204	-0,087	-0,077
El Salvador	0,164	2,704	1,724	-0,496	-1,325	-0,496	-0,166	-0,166	-0,506	0,927	0,167 ^d	-0,480	-1,072	-0,806	-0,493	-0,656
Panamá	0,455	1,849	1,449	0,055	-1,563	0,169	0,255	0,312	-0,052 ^b	0,634	0,481	-0,192	-0,989	-0,166	-0,122	-0,109
Paraguay	-0,482	1,806	1,118	-0,941	-1,696	-0,594	-0,712	-0,538	-0,418	0,518	0,276 ^d	-0,618	-0,657	-0,529	-0,518	-0,474
Perú	0,182	1,865	1,371	-0,394	-1,622	-0,194	-0,106	-0,006 ^c	-0,029 ^c	0,765	0,800	-0,282	-1,647	-0,147	-0,156	-0,088
República Dominicana	1,206	3,400	2,724	0,865	-0,616	0,924	1,035	1,065	0,341	1,482	1,188	0,224	-0,196	0,188 ^b	0,282	0,265
Uruguay	-0,013 ^d	0,993	0,773	-0,327	-1,553	-0,247	-0,170	-0,130	-0,080	0,060	0,020	-0,120	-0,102	-0,120	-0,100	-0,100

Fuente: Elaboración propia.

^a Se resaltan los valores de aquellos modelos que reportan el menor sesgo.

^b p<0.10.

^c p<0.05.

^d p<0.01.

Cuadro 14
La raíz del error cuadrático medio de pronóstico de los distintos modelos (2003-2019)^a

País	Pobreza								Pobreza extrema							
	CEPAL	PVAR	PVAR+ exog	LASSO	LASSO agregado	LASSO+ exog	Pooling	Pooling	CEPAL	PVAR	PVAR+ exog	LASSO	LASSO agregado	LASSO+ exog	Pooling	Pooling
	(0)	(1)	(2)	(3)	(4)	(5)	(0)+(3)	(0)+(5)	(0)	(1)	(2)	(3)	(4)	(5)	(0)+(3)	(0)+(5)
Argentina	4,162	4,381	3,945	4,662	6,863	3,525^b	4,363	3,804	2,869	2,798	2,709	3,017	3,367	2,728	2,942	2,797
Bolivia (Estado Plurinacional de)	3,413	3,536	3,299	3,496	7,64	3,361	3,425	3,355	4,521	4,122	4,053	4,590	5,267	4,538	4,548	4,523
Brasil	1,337	2,437	1,998	1,539	2,177	1,362	1,395	1,272	0,563	0,823	0,753	0,553	1,024	0,533	0,553	0,535
Colombia	1,153	2,742	2,100	1,287	2,531	1,368	1,180	1,186	0,923	1,472	1,131	1,034	1,869	1,099	0,946	0,971
Costa Rica	1,996	2,964	2,574	1,890	1,884	2,031	1,915	1,977	0,692	0,802	0,750	0,701	0,670	0,715	0,692	0,698
Ecuador	1,581	3,311	2,711	2,095	3,023	2,111	1,758	1,763	1,092	1,436	1,272	1,195	1,572	1,221	1,121	1,134
El Salvador	1,882	3,464	2,678	1,932	2,707	2,015	1,867	1,902	1,693	2,298	1,560	1,889	2,089	1,776	1,693	1,724
Panamá	1,158	2,263	1,932	1,125	2,272	1,260	1,071	1,150	1,116	1,264	1,192	1,072	1,413	1,082	1,075	1,087
Paraguay	3,151	3,378	2,999	3,222	3,891	3,230	3,153	3,139	2,732	2,672	2,622	2,805	3,114	2,844	2,757	2,778
Perú	7,034	7,502	7,560	7,413	6,606	7,105	7,214	7,060	4,120	4,093	4,289	4,366	4,914	4,133	4,240	4,125
República Dominicana	3,587	5,078	4,705	3,841	5,797	3,773	3,698	3,650	1,615	2,427	2,153	1,778	3,175	1,601	1,690	1,587
Uruguay	1,909	2,317	2,288	2,053	4,274	1,931	1,966	1,895	0,527	0,517	0,542	0,581	1,372	0,582	0,552	0,552

Fuente: Elaboración propia.

^aSe resaltan los valores de aquellos modelos que reportan la menor raíz del error cuadrático medio.

^bp<0.05.

Cuadro 15
El error medio absoluto de pronóstico de los distintos modelos (2003-2019)^a

País	Pobreza								Pobreza extrema							
	CEPAL	PVAR	PVAR+ exog	LASSO	LASSO agregado	LASSO+ exog	Pooling	Pooling	CEPAL	PVAR	PVAR+ exog	LASSO	LASSO agregado	LASSO+ exog	Pooling	Pooling
	(0)	(1)	(2)	(3)	(4)	(5)	(0)+(3)	(0)+(5)	(0)	(1)	(2)	(3)	(4)	(5)	(0)+(3)	(0)+(5)
Argentina	3,140	3,387	3,107	3,113	4,692	2,507^b	3,080	2,810	1,387	1,347	1,320	1,427	1,853	1,340	1,400	1,357
Bolivia (Estado Plurinacional de)	2,329	2,775	2,502	2,222	5,380	2,082	2,249	2,152	2,895	2,581	2,575	2,988	3,824	2,928	2,941	2,911
Brasil	1,001	2,095	1,558	1,224	1,879	1,067	1,033	0,920	0,499	0,692	0,611	0,492	0,734	0,442	0,495	0,458
Colombia	0,910	2,530	1,843	1,010	1,957	1,090	0,953	0,920	0,727	1,260	0,893	0,793	1,666	0,887	0,747	0,773
Costa Rica	1,476	2,529	2,129	1,424	1,576	1,553	1,403	1,468	0,571	0,653	0,618	0,576	0,563	0,606	0,568	0,576
Ecuador	1,267	2,833	2,313	1,801	2,293	1,754	1,507	1,484	0,903	1,169	1,036	0,997	1,181	1,031	0,914	0,931
El Salvador	1,657	3,050	2,270	1,677	2,266	1,810	1,627	1,694	1,352	1,505	1,060	1,512	1,688	1,439	1,332	1,382
Panamá	0,930	1,915	1,595	0,917	1,878	1,110	0,870	0,973	0,801	1,021	0,908	0,857	1,260	0,817	0,827	0,807
Paraguay	2,188	2,571	2,388	2,506	3,220	2,535	2,347	2,362	2,006	2,000	1,994	2,124	2,442	2,165	2,065	2,079
Perú	3,712	4,159	4,112	3,818	4,415	3,582	3,765	3,647	1,982	1,988	2,200	2,106	3,043	2,029	2,044	2,000
República Dominicana	2,488	3,682	3,359	2,688	3,994	2,794	2,547	2,618	1,106	1,600	1,341	1,224	2,370	1,200	1,165	1,112
Uruguay	1,173	1,340	1,187	1,300	3,327	1,180	1,203	1,137	0,267	0,247	0,233	0,293	1,199	0,307	0,280	0,287

Fuente: Elaboración propia.

^aSe resaltan los valores de aquellos modelos que reportan el menor error medio absoluto.

^bp<0.05.

Nuevamente, si se trabaja con las proyecciones de CEPAL o alguna combinación con LASSO habría ganancias predictivas en 9 de los 12 países analizados para la tasa de pobreza, mientras que, para la tasa de pobreza extrema, las proyecciones de CEPAL o los modelos que incorporan las variables exógenas parecen ser mejores en la mitad de los casos. Sin embargo, no se observan ganancias predictivas cuando se realizan combinaciones de ambas alternativas.

Por último, al trabajar con una función de pérdida absoluta pero agregada en función de la mediana como en el cuadro 16, en lugar de la media, tenemos una medida más robusta a la presencia de valores atípicos en las predicciones de pobreza.

Si bien no surgen diferencias estadísticamente significativas, con la excepción nuevamente de Argentina (cuando se trabaja con modelos que consideren variables macroeconómicas exógenas), se evidencia una mayor paridad en el desempeño relativo entre las proyecciones de CEPAL y las proyecciones realizadas por LASSO (o alguna combinación entre ambos).

Por otro lado, se presentan las medidas de desempeño (error medio, la raíz del error cuadrático medio, el error medio absoluto y el error mediano absoluto) sobre los pronósticos de las tasas de pobreza y pobreza extrema de tres grupos de interés: las mujeres, los desocupados, y los jóvenes (entre 15 y 24 años).

Se utilizan el modelo de CEPAL (modelo 0) y los 4 modelos descritos anteriormente junto con las dos combinaciones (*pooling*) propuestas. No se incorpora el modelo 5 por no disponer de datos de ingreso per cápita y Gini a nivel de subgrupos de interés.

Cuadro 16
El error mediano absoluto del pronóstico de los distintos modelos (2003-2019)^a

País	Pobreza								Pobreza extrema							
	CEPAL	PVAR	PVAR+ exog	LASSO	LASSO agregado	LASSO+ exog	Pooling	Pooling	CEPAL	PVAR	PVAR+ exog	LASSO	LASSO agregado	LASSO+ exog	Pooling	Pooling
	(0)	(1)	(2)	(3)	(4)	(5)	(0)+(3)	(0)+(5)	(0)	(1)	(2)	(3)	(4)	(5)	(0)+(3)	(0)+(5)
Argentina	2,300	2,600	2,500	2,500	3,411	1,700	2,350	2,050	0,600	0,700	0,700	0,600	0,737	0,600	0,550	0,600
Bolivia (Estado Plurinacional de)	2,000	2,800	2,000	1,500	2,904	1,600	1,600	1,750	1,600	1,571	1,671	1,600	1,892	1,500	1,450	1,550
Brasil	0,800	1,900	1,100	0,989	1,860	1,000	0,700	0,525	0,500	0,700	0,638	0,500	0,440	0,450	0,488	0,450
Colombia	0,700	2,800	2,000	0,700	1,595	1,000	0,700	0,750	0,600	1,000	0,800	0,400	1,559	0,800	0,450	0,600
Costa Rica	1,000	2,100	1,700	0,900	1,466	1,000	0,950	1,000	0,600	0,600	0,500	0,600	0,573	0,500	0,600	0,550
Ecuador	1,100	2,900	2,400	1,800	1,649	2,100	1,361	1,350	0,900	1,000	0,740	1,000	0,695	0,900	0,850	0,900
El Salvador	1,600	2,886	1,886	1,700	1,740	1,786	1,550	1,450	1,300	0,800	0,600	1,200	1,254	1,100	1,043	1,000
Panamá	0,600	1,500	1,500	0,600	1,316	1,000	0,810	0,860	0,487	0,800	0,700	0,600	1,334	0,600	0,550	0,500
Paraguay	1,100	2,100	2,300	1,700	2,482	2,100	1,500	1,650	1,400	1,400	1,400	1,500	1,799	1,800	1,500	1,650
Perú	1,200	1,800	1,700	0,900	2,297	0,700	1,100	0,850	0,600	0,700	0,600	0,500	1,741	0,500	0,500	0,550
República Dominicana	1,900	2,400	1,900	1,900	2,629	1,600	2,050	1,900	0,600	0,700	0,900	0,700	2,182	0,800	0,650	0,700
Uruguay	0,600	0,700	0,600	0,700	2,852	0,600	0,700	0,650	0,100	0,100	0,000	0,100	1,116	0,100	0,100	0,100

Fuente: Elaboración propia.

^aSe resaltan los valores de aquellos modelos que reportan el menor error mediano absoluto.

A. Pronósticos de pobreza en mujeres

A partir de los datos de pobreza en mujeres para los países de la región provistos por CEPAL, se evaluó el desempeño de los pronósticos de cada uno de los modelos (reportados en los cuadros 17 a 20) entre 2003 y 2019.

En algunos países el modelo de CEPAL tiende a sobrepronosticar las tasas de pobreza en las mujeres y en otros a subpronosticar. De todas formas, no es el modelo con menor sesgo, otras estrategias alternativas como el PVAR y LASSO, en algunos casos, reportan un menor sesgo (cuadro 17).

Cuando trabajamos con una función de pérdida cuadrática, el modelo con mejor desempeño es muy específico a cada país. El modelo de CEPAL parece dominar en los pronósticos de pobreza en mujeres para Colombia, Panamá y República Dominicana, aunque no tienen diferencias estadísticamente significativas con el segundo mejor modelo (cuadro 18).

Cuando trabajamos con una función de pérdida absoluta, como el error medio absoluto, el PVAR (incluyendo o no el ingreso cuadrático en forma exógena) parece ser el modelo con mejor desempeño para la región, seguido por una selección a través de LASSO (cuadro 19).

Por último, cuando evaluamos en función de una medida más robusta a la presencia de valores extremos, como el error mediano absoluto (MdAE), no aparecen diferencias estadísticamente significativas entre los distintos modelos (cuadro 20).

En resumen, no hay un modelo que claramente pronostique mejor la tasa de pobreza en las mujeres en los distintos países, ni tampoco existe una combinación de pronósticos que resulte superadora. Las ganancias (marginales) son muy específicas a cada país.

B. Pronósticos de pobreza en desocupados

A partir de los datos de pobreza en desocupados para los países de la región provistos por CEPAL, se evaluó el desempeño de los pronósticos de cada uno de los modelos (reportados en los cuadros 21 a 24) entre 2003 y 2019.

Al analizar los sesgos que cometen los distintos modelos de pronósticos de las tasas de pobreza de los desocupados, se observa que todos tienden a sobrepronosticar dichas tasas. El modelo que sistemáticamente comete un menor sesgo es el PVAR (sin variables exógenas ni restricciones). De todas formas, el sesgo es importante, entre 8.6 y 20.1 puntos porcentuales (cuadro 21).

En función de la raíz del error cuadrático medio (RMSE), se observa que el modelo con mejor desempeño para pronosticar las tasas de pobreza de los desocupados en la región es el PVAR (sin variables exógenas ni restricciones), aunque las diferencias no son estadísticamente significativas respecto del segundo mejor modelo que suele ser el PVAR que incluye variables exógenas, pero claramente representa una mejor estrategia que usar el modelo base de CEPAL (cuadro 22).

Al cambiar de función de pérdida y evaluar en función del error medio absoluto (MAE), los resultados del cuadro 23 anterior se mantienen: el modelo VAR en panel sigue siendo la mejor estrategia de pronósticos de las tasas de pobreza en desocupados de la región.

Resultados similares se observan cuando se evalúan los pronósticos en función del error mediano absoluto (MdAE) aunque con algunas excepciones (cuadro 24). El modelo de CEPAL muestra ganancias solo para pronosticar la tasa de pobreza extrema de los desocupados en Bolivia, y el modelo VAR en panel que incluye al ingreso cuadrático en forma exógena muestra un desempeño muy similar al PVAR.

Cuadro 17
El error medio de pronóstico de pobreza en mujeres^a

País	Error medio													
	Pobreza							Pobreza extrema						
	CEPAL	PVAR	PVAR+ exog	LASSO	LASSO+ exog.	Pooling	Pooling	CEPAL	PVAR	PVAR+ exog	LASSO	LASSO+ exog.	Pooling	Pooling
(0)	(1)	(2)	(3)	(5)	(0)+(3)	(0)+(5)	(0)	(1)	(2)	(3)	(5)	(0)+(3)	(0)+(5)	
Argentina	-2,887	-1,420	-1,713	-3,847	-3,540	-3,367	-3,213	-2,553	-2,333	-2,347	-2,727	-2,653	-2,640	-2,603
Bolivia (Estado Plurinacional de)	3,362	5,185	4,492	2,554	2,900	2,958	3,131	3,554	4,554	4,177	3,115	3,208	3,335	3,381
Brasil	-1,733	0,793	0,127	-1,900	-2,013	-1,817	-1,873	-0,500	0,080	-0,067	-0,527	-0,560	-0,513	-0,530
Colombia	-1,050	1,386	0,686	-1,464	-1,543	-1,257	-1,296	-1,164	0,164	-0,250	-1,486	-1,464	-1,325	-1,314
Costa Rica	-1,365	0,282	-0,229	-1,876	-1,800	-1,621	-1,582	-0,953	-0,418	-0,612	-1,112	-1,059	-1,032	-1,006
Ecuador	1,000	3,379	2,657	0,357	0,350	0,679	0,675	0,407	1,386	1,114	0,186	0,186	0,296	0,296
El Salvador	1,046	3,915	3,000	0,469	0,385	0,758	0,715	1,623	2,715	2,408	1,354	1,292	1,488	1,458
Panamá	-0,746	0,808	0,323	-1,077	-0,985	-0,912	-0,865	-0,185	0,538	0,377	-0,315	-0,277	-0,250	-0,231
Paraguay	-0,982	1,506	0,794	-1,412	-0,965	-1,197	-0,974	0,053	1,118	0,876	-0,059	0,029	-0,003	0,041
Perú	4,718	6,324	5,876	4,159	4,353	4,438	4,535	2,847	3,482	3,306	2,629	2,759	2,738	2,803
República Dominicana	-1,253	0,982	0,288	-1,588	-1,565	-1,421	-1,409	-1,059	0,147	-0,194	-1,224	-1,212	-1,141	-1,135
Uruguay	-1,420	-0,360	-0,593	-1,733	-1,640	-1,577	-1,530	-0,540	-0,367	-0,433	-0,593	-0,580	-0,567	-0,560

Fuente: Elaboración propia.

^aSe resaltan los valores de aquellos modelos que reportan el menor error medio.

Cuadro 18
Raíz del error cuadrático medio del pronóstico de pobreza en mujeres^a

País	Pobreza						Pobreza extrema							
	CEPAL	PVAR	PVAR+ exog	LASSO	LASSO+ exog	Pooling	Pooling	CEPAL	PVAR	PVAR+ exog	LASSO	LASSO+ exog	Pooling	Pooling
	(0)	(1)	(2)	(3)	(5)	(0)+(3)	(0)+(5)	(0)	(1)	(2)	(3)	(5)	(0)+(3)	(0)+(5)
Argentina	5,302	4,384	4,304	5,966	5,245	5,603	5,236	3,917	3,756	3,695	4,075	3,851	3,994	3,882
Bolivia (Estado Plurinacional de)	5,377	6,345	5,900	4,828	4,941	5,091	5,143	5,267	6,185	5,737	4,935	4,990	5,096	5,124
Brasil	2,106	1,765	1,571	2,285	2,316	2,181	2,195	0,846	0,722	0,748	0,840	0,863	0,840	0,850
Colombia	1,722	2,123	1,732	2,032	2,259	1,867	1,959	1,386	0,936	0,965	1,753	1,769	1,562	1,557
Costa Rica	2,556	2,029	2,033	2,956	2,951	2,742	2,736	1,440	1,068	1,203	1,593	1,557	1,514	1,495
Ecuador	1,749	3,906	3,272	1,807	1,959	1,679	1,737	1,306	1,729	1,562	1,263	1,320	1,266	1,293
El Salvador	1,918	4,300	3,457	1,733	1,672	1,795	1,747	2,284	3,191	2,915	2,088	2,023	2,182	2,148
Panamá	1,622	1,906	1,728	1,887	1,947	1,721	1,751	1,625	1,811	1,767	1,597	1,641	1,604	1,624
Paraguay	3,626	3,437	3,321	3,838	3,565	3,704	3,537	2,374	2,623	2,453	2,454	2,379	2,404	2,363
Perú	8,825	10,020	9,848	8,905	8,764	8,854	8,786	5,238	5,655	5,576	5,236	5,210	5,234	5,223
República Dominicana	3,600	4,026	3,953	4,018	3,811	3,798	3,679	2,415	2,298	2,185	2,537	2,289	2,470	2,338
Uruguay	2,773	2,478	2,576	3,002	2,836	2,880	2,790	0,985	0,738	0,835	1,074	1,064	1,029	1,023

Fuente: Elaboración propia.

^aSe resaltan los valores de aquellos modelos que reportan la menor raíz del error cuadrático medio.

Cuadro 19
Error medio absoluto del pronóstico de pobreza en mujeres^a

País	Pobreza							Pobreza extrema						
	CEPAL	PVAR	PVAR+ exog	LASSO	LASSO+ exog	Pooling	Pooling	CEPAL	PVAR	PVAR+ exog	LASSO	LASSO+ exog	Pooling	Pooling
	(0)	(1)	(2)	(3)	(5)	(0)+(3)	(0)+(5)	(0)	(1)	(2)	(3)	(5)	(0)+(3)	(0)+(5)
Argentina	4,127	3,313	3,247	4,660	4,247	4,260	4,040	2,567	2,373	2,387	2,767	2,653	2,667	2,603
Bolivia (Estado Plurinacional de)	4,854	5,923	5,446	4,169	4,269	4,512	4,531	4,400	5,185	4,823	3,992	4,054	4,181	4,227
Brasil	1,907	1,407	1,153	2,060	2,053	1,983	1,973	0,660	0,547	0,573	0,673	0,667	0,667	0,663
Colombia	1,507	1,643	1,357	1,707	1,757	1,607	1,632	1,279	0,779	0,836	1,586	1,579	1,432	1,429
Costa Rica	1,718	1,529	1,488	2,100	2,141	1,856	1,888	1,082	0,782	0,906	1,206	1,188	1,144	1,135
Ecuador	1,471	3,407	2,829	1,443	1,593	1,386	1,454	1,007	1,486	1,286	0,971	1,043	0,989	1,025
El Salvador	1,615	3,946	3,154	1,438	1,400	1,519	1,485	1,715	2,715	2,454	1,477	1,400	1,596	1,558
Panamá	1,454	1,608	1,477	1,615	1,585	1,527	1,512	1,231	1,323	1,285	1,208	1,231	1,212	1,231
Paraguay	2,841	2,906	2,841	3,024	2,882	2,874	2,791	1,818	1,647	1,571	1,859	1,806	1,809	1,782
Perú	6,094	7,406	7,076	5,876	5,906	5,926	5,941	3,329	3,729	3,612	3,253	3,253	3,291	3,279
República Dominicana	3,018	2,829	2,982	3,471	3,318	3,238	3,144	2,024	1,888	1,829	2,129	1,918	2,071	1,959
Uruguay	2,300	1,787	1,913	2,560	2,360	2,430	2,330	0,553	0,393	0,447	0,593	0,593	0,573	0,573

Fuente: Elaboración propia.

^aSe resaltan los valores de aquellos modelos que reportan el menor error medio absoluto.

Cuadro 20
Error mediano absoluto del pronóstico de pobreza en mujeres^a

País	Pobreza							Pobreza extrema						
	CEPAL	PVAR	PVAR+ exog	LASSO	LASSO+ exog	Pooling	Pooling	CEPAL	PVAR	PVAR+ exog	LASSO	LASSO+ exog	Pooling	Pooling
	(0)	(1)	(2)	(3)	(5)	(0)+(3)	(0)+(5)	(0)	(1)	(2)	(3)	(5)	(0)+(3)	(0)+(5)
Argentina	3,800	2,300	2,500	4,700	4,300	4,250	4,050	1,400	1,200	1,200	1,700	1,600	1,550	1,450
Bolivia (Estado Plurinacional de)	5,000	6,200	6,100	3,900	4,100	4,600	4,650	4,600	5,200	5,000	4,400	4,500	4,550	4,600
Brasil	2,000	1,100	0,800	2,100	2,200	2,000	2,000	0,500	0,400	0,400	0,500	0,600	0,500	0,450
Colombia	1,200	1,400	0,950	1,250	1,350	1,100	1,100	1,400	0,650	0,800	1,500	1,500	1,450	1,550
Costa Rica	0,900	1,600	1,300	1,300	1,100	1,100	1,000	0,800	0,500	0,600	0,900	0,900	0,850	0,850
Ecuador	1,300	3,300	2,650	1,350	1,150	1,075	1,200	0,800	1,450	1,250	0,800	0,950	0,800	0,850
El Salvador	1,400	4,400	3,300	1,300	1,500	1,050	1,250	1,400	2,800	2,300	1,200	1,100	1,300	1,250
Panamá	1,100	1,200	1,100	1,300	1,300	1,250	1,150	0,800	1,000	0,800	0,500	0,500	0,700	0,600
Paraguay	2,100	2,900	2,500	2,400	2,500	2,400	2,100	1,300	1,200	1,000	1,400	1,200	1,350	1,200
Perú	4,300	5,300	5,200	3,800	3,700	4,150	4,050	1,900	2,300	2,500	1,800	1,900	1,850	1,900
República Dominicana	2,800	2,000	2,200	3,700	3,100	3,200	2,850	1,500	1,500	1,600	1,600	1,300	1,400	1,300
Uruguay	1,900	0,900	1,200	2,000	2,000	1,950	1,850	0,200	0,100	0,100	0,200	0,200	0,200	0,200

Fuente: Elaboración propia.

^aSe resaltan los valores de aquellos modelos que reportan el menor error mediano absoluto.

Cuadro 21
El error medio de pronóstico de pobreza en desocupados^a

País	Error medio													
	Pobreza							Pobreza extrema						
	CEPAL	PVAR	PVAR+ exog	LASSO	LASSO+ exog	Pooling	Pooling	CEPAL	PVAR	PVAR+ exog	LASSO	LASSO+ exog	Pooling	Pooling
(0)	(1)	(2)	(3)	(5)	(0)+(3)	(0)+(5)	(0)	(1)	(2)	(3)	(5)	(0)+(3)	(0)+(5)	
Argentina	-16,360	-14,427	-14,880	-17,213	-17,107	-16,787	-16,733	-12,493	-12,207	-12,053	-13,167	-12,840	-12,830	-12,667
Bolivia (Estado Plurinacional de)	-16,177	-14,031	-14,623	-16,954	-16,385	-16,565	-16,281	-14,154	-14,138	-14,477	-15,662	-15,192	-14,908	-14,673
Brasil	-21,053	-18,153	-18,960	-21,320	-21,653	-21,187	-21,353	-16,027	-14,713	-15,033	-16,127	-16,493	-16,077	-16,260
Colombia	-12,114	-9,836	-10,507	-12,564	-12,621	-12,339	-12,368	-12,350	-10,671	-11,193	-12,771	-12,736	-12,561	-12,543
Costa Rica	-22,124	-20,094	-20,688	-22,600	-22,576	-22,362	-22,350	-18,824	-17,241	-18,047	-19,359	-19,212	-19,091	-19,018
Ecuador	-14,257	-12,036	-12,629	-14,707	-14,814	-14,482	-14,536	-10,300	-8,950	-9,443	-10,643	-10,807	-10,471	-10,554
El Salvador	-13,185	-10,662	-11,346	-13,815	-13,669	-13,500	-13,427	-13,023	-11,146	-11,746	-13,492	-13,600	-13,258	-13,312
Panamá	-17,615	-15,254	-16,400	-18,208	-18,108	-17,912	-17,862	-12,408	-11,638	-11,962	-12,600	-12,423	-12,504	-12,415
Paraguay	-13,947	-10,806	-11,994	-14,576	-14,476	-14,262	-14,212	-9,018	-7,712	-8,212	-9,418	-9,094	-9,218	-9,056
Perú	-10,312	-8,576	-9,088	-10,700	-10,606	-10,506	-10,459	-7,529	-6,682	-7,053	-7,806	-7,794	-7,668	-7,662
República Dominicana	-16,812	-15,247	-15,665	-17,112	-17,012	-16,962	-16,912	-16,182	-14,147	-14,335	-16,341	-16,429	-16,262	-16,306
Uruguay	-12,000	-10,053	-10,540	-12,700	-12,327	-12,350	-12,163	-4,433	-3,893	-3,987	-4,673	-4,640	-4,553	-4,537

Fuente: Elaboración propia.

^aSe resaltan los valores de aquellos modelos que reportan el menor error medio.

Cuadro 22
Raíz del error cuadrático medio del pronóstico de pobreza en desocupados^a

País	RMSE													
	Pobreza							Pobreza extrema						
	CEPAL	PVAR	PVAR+ exog	LASSO	LASSO+ exog	Pooling	Pooling	CEPAL	PVAR	PVAR+ exog	LASSO	LASSO+ exog	Pooling	Pooling
(0)	(1)	(2)	(3)	(5)	(0)+(3)	(0)+(5)	(0)	(1)	(2)	(3)	(5)	(0)+(3)	(0)+(5)	
Argentina	17,215	15,280	15,760	18,002	17,936	17,599	17,558	13,230	12,878	12,685	13,797	13,387	13,505	13,296
Bolivia (Estado Plurinacional de)	18,187	16,318	16,755	18,858	18,177	18,515	18,170	17,036	16,911	17,299	18,283	17,634	17,571	17,245
Brasil	21,439	18,563	19,405	21,730	21,954	21,582	21,690	16,530	15,274	15,602	16,648	16,936	16,588	16,722
Colombia	12,192	9,947	10,608	12,639	12,696	12,414	12,438	12,499	10,790	11,326	12,928	12,872	12,711	12,681
Costa Rica	22,720	20,792	21,405	23,171	23,130	22,943	22,922	19,251	17,707	18,545	19,852	19,674	19,549	19,460
Ecuador	15,793	13,793	14,281	16,230	16,329	16,000	16,053	11,820	10,425	10,979	12,089	12,280	11,951	12,045
El Salvador	13,804	11,671	12,276	14,302	14,256	14,043	14,028	14,531	13,141	13,630	14,880	14,954	14,703	14,739
Panamá	18,588	16,362	17,406	19,057	18,890	18,816	18,731	13,764	13,020	13,336	13,957	13,779	13,859	13,769
Paraguay	17,001	14,625	15,296	17,215	17,168	17,092	17,060	11,519	10,084	10,456	11,620	11,195	11,556	11,333
Perú	13,597	12,732	13,115	14,049	13,949	13,817	13,765	8,657	7,846	8,209	8,942	8,884	8,797	8,767
República Dominicana	17,454	15,943	16,384	17,761	17,636	17,601	17,538	17,412	15,620	15,790	17,549	17,549	17,478	17,475
Uruguay	13,041	11,111	11,589	13,731	13,405	13,379	13,210	7,085	6,518	6,604	7,337	7,324	7,209	7,203

Fuente: Elaboración propia.

^a Se resaltan los valores de aquellos modelos que reportan la menor raíz del error cuadrático medio.

Cuadro 23
Error medio absoluto del pronóstico de pobreza en desocupados^a

País	Pobreza							Pobreza extrema						
	CEPAL	PVAR	PVAR+ exog	LASSO	LASSO+ exog	Pooling	Pooling	CEPAL	PVAR	PVAR+ exog	LASSO	LASSO+ exog	Pooling	Pooling
	(0)	(1)	(2)	(3)	(5)	(0)+(3)	(0)+(5)	(0)	(1)	(2)	(3)	(5)	(0)+(3)	(0)+(5)
Argentina	16,360	14,427	14,880	17,213	17,107	16,787	16,733	12,493	12,207	12,053	13,167	12,840	12,830	12,667
Bolivia (Estado Plurinacional de)	17,238	15,138	15,731	17,908	17,338	17,573	17,288	14,600	14,738	15,077	16,108	15,638	15,354	15,119
Brasil	21,053	18,153	18,960	21,320	21,653	21,187	21,353	16,027	14,713	15,033	16,127	16,493	16,077	16,260
Colombia	12,114	9,836	10,507	12,564	12,621	12,339	12,368	12,350	10,671	11,193	12,771	12,736	12,561	12,543
Costa Rica	22,124	20,094	20,688	22,600	22,576	22,362	22,350	18,824	17,241	18,047	19,359	19,212	19,091	19,018
Ecuador	14,271	12,250	12,829	14,793	14,900	14,532	14,586	10,357	9,021	9,500	10,657	10,821	10,507	10,589
El Salvador	13,185	10,662	11,346	13,815	13,669	13,500	13,427	13,023	11,146	11,746	13,492	13,600	13,258	13,312
Panamá	17,615	15,254	16,400	18,208	18,108	17,912	17,862	12,408	11,638	11,962	12,600	12,423	12,504	12,415
Paraguay	15,065	12,441	13,535	15,447	15,347	15,256	15,206	9,547	8,135	8,635	9,747	9,424	9,547	9,385
Perú	12,324	11,141	11,594	12,771	12,676	12,547	12,500	7,941	7,235	7,594	8,218	8,206	8,079	8,074
República Dominicana	16,812	15,247	15,665	17,112	17,012	16,962	16,912	16,182	14,147	14,335	16,341	16,429	16,262	16,306
Uruguay	12,000	10,053	10,540	12,700	12,327	12,350	12,163	4,460	3,947	4,053	4,700	4,667	4,580	4,563

Fuente: Elaboración propia.

^a Se resaltan los valores de aquellos modelos que reportan el menor error medio absoluto.

Cuadro 24
Error mediano absoluto del pronóstico de pobreza en desocupados^a

País	Pobreza							Pobreza extrema						
	CEPAL	PVAR	PVAR+ exog	LASSO	LASSO+ exog	Pooling	Pooling	CEPAL	PVAR	PVAR+ exog	LASSO	LASSO+ exog	Pooling	Pooling
	(0)	(1)	(2)	(3)	(5)	(0)+(3)	(0)+(5)	(0)	(1)	(2)	(3)	(5)	(0)+(3)	(0)+(5)
Argentina	17,000	15,200	15,200	17,000	17,700	17,000	17,000	10,800	10,500	10,500	11,000	11,000	10,800	10,900
Bolivia (Estado Plurinacional de)	16,000	15,400	16,000	17,700	17,100	16,850	16,600	13,300	13,900	13,900	14,800	14,800	13,300	13,300
Brasil	22,500	19,500	20,100	23,200	23,400	22,900	22,750	15,300	14,600	14,600	15,300	15,100	15,300	15,200
Colombia	12,050	9,350	10,400	12,400	12,650	12,225	12,425	11,900	10,200	10,850	12,200	13,000	12,050	12,375
Costa Rica	21,300	19,700	19,700	22,800	21,900	21,900	21,450	18,700	17,800	18,500	19,000	19,000	18,850	18,850
Ecuador	14,700	11,950	12,900	15,050	15,400	14,925	15,375	9,300	8,800	9,050	10,700	10,700	10,000	10,000
El Salvador	13,600	11,800	12,600	13,900	13,700	13,750	13,700	11,300	9,400	9,800	11,800	12,000	11,550	11,550
Panamá	17,900	14,800	15,200	17,500	17,300	17,700	17,600	10,400	10,300	10,300	10,400	10,400	10,400	10,400
Paraguay	14,200	11,600	12,600	14,200	16,000	14,200	15,650	8,500	6,500	8,000	8,600	8,600	8,550	8,550
Perú	11,500	9,400	9,800	11,500	12,000	11,500	11,500	7,600	6,900	7,600	8,100	8,100	8,100	8,100
República Dominicana	15,400	14,000	14,000	16,800	16,800	15,400	15,400	16,500	14,800	14,800	16,500	16,500	16,500	16,500
Uruguay	11,600	10,300	9,800	12,500	11,400	12,050	11,500	1,800	1,300	1,500	1,800	1,800	1,800	1,800

Fuente: Elaboración propia.

^a Se resaltan los valores de aquellos modelos que reportan el menor error mediano absoluto.

C. Pronósticos de pobreza en jóvenes

A partir de los datos de pobreza en jóvenes entre 15 y 24 años para los países de la región provistos por CEPAL, se evaluó el desempeño de los pronósticos de cada uno de los modelos (reportados en los cuadros 25 a 28) entre 2003 y 2019.

Al evaluar el sesgo que cometen los distintos modelos de pronóstico observamos un resultado similar al caso de las tasas de pobreza en los desocupados, es decir, los distintos modelos tienden a sobrepronosticar dichas tasas, aunque la magnitud del sesgo es menor. El VAR en panel es el modelo que comete menores sesgos en la región, a excepción del modelo de CEPAL para pronosticar la tasa de pobreza de los jóvenes en Bolivia (cuadro 25).

Según la raíz del error cuadrático medio (RMSE) (cuadro 26), el modelo PVAR nuevamente parece ser el ganador para pronosticar la tasa de pobreza de los jóvenes en la región. Cuando se evalúa el desempeño de los distintos modelos para pronosticar la tasa de pobreza extrema, LASSO aparece como una posible alternativa en Bolivia y Ecuador.

El error medio absoluto (MAE) reporta resultados similares al RMSE (cuadro 27). El PVAR parece ser el mejor modelo en términos de pronósticos de la tasa de pobreza para los jóvenes de la región, a excepción de LASSO para Bolivia.

Según el error mediano absoluto (MdAE), el modelo base de CEPAL tiene un mejor desempeño de pronóstico que el resto de las alternativas únicamente para las tasas de pobreza de jóvenes en Ecuador, no así para las tasas de pobreza extrema (cuadro 28). El PVAR nuevamente parece ser una estrategia mejor en términos de pronóstico tanto de la tasa de pobreza como pobreza extrema de los jóvenes en la región.

En general, se observan para todos los modelos sesgos muy importantes en las proyecciones de las tasas de pobreza en los jóvenes. Sin embargo, el VAR en panel (en algunos casos, incluyendo variables exógenas) parece ser el modelo con mejor desempeño a la hora de proyectar la pobreza en este grupo, al igual que cuando se proyectaba la pobreza en los desocupados.

Cuadro 25
El error medio de pronóstico de pobreza en jóvenes^a

País	Error medio													
	Pobreza							Pobreza extrema						
	CEPAL	PVAR	PVAR+ exog	LASSO	LASSO+ exog	Pooling	Pooling	CEPAL	PVAR	PVAR+ exog	LASSO	LASSO+ exog	Pooling	Pooling
(0)	(1)	(2)	(3)	(5)	(0)+(3)	(0)+(5)	(0)	(1)	(2)	(3)	(5)	(0)+(3)	(0)+(5)	
Argentina	-5,573	-3,493	-3,960	-6,633	-6,227	-6,103	-5,900	-4,580	-4,413	-4,487	-4,827	-4,713	-4,703	-4,647
Bolivia (Estado Plurinacional de)	-0,769	1,585	0,762	-1,769	-1,585	-1,269	-1,177	1,405	2,223	1,954	1,023	1,062	1,214	1,233
Brasil	-11,007	-7,733	-8,587	-11,260	-11,313	-11,133	-11,160	-4,253	-3,387	-3,587	-4,327	-4,393	-4,290	-4,323
Colombia	-4,143	-1,214	-2,064	-4,671	-4,736	-4,407	-4,439	-3,186	-1,807	-2,207	-3,571	-3,479	-3,379	-3,332
Costa Rica	-7,453	-5,388	-6,006	-7,929	-7,712	-7,691	-7,582	-2,747	-2,324	-2,506	-2,894	-2,935	-2,821	-2,841
Ecuador	-2,543	0,221	-0,671	-3,300	-3,164	-2,921	-2,854	0,171	0,993	0,664	-0,314	-0,436	-0,071	-0,132
El Salvador	-7,638	-4,838	-5,777	-8,346	-8,508	-7,992	-8,073	-2,985	-1,585	-2,138	-3,369	-3,338	-3,177	-3,162
Panamá	-3,900	-1,685	-2,523	-4,115	-4,046	-4,008	-3,973	-2,646	-1,900	-2,115	-2,923	-2,769	-2,785	-2,708
Paraguay	-2,524	-0,171	-0,218	-2,876	-2,847	-2,700	-2,685	-0,524	0,341	0,176	-0,729	-0,606	-0,626	-0,565
Perú	-3,647	-1,706	-2,359	-4,165	-4,006	-3,906	-3,826	-2,271	-1,453	-1,841	-2,600	-2,453	-2,435	-2,362
República Dominicana	-3,118	-0,535	-1,329	-3,518	-3,488	-3,318	-3,303	-1,947	-0,841	-1,171	-2,012	-2,118	-1,979	-2,032
Uruguay	-6,820	-5,407	-5,640	-7,207	-7,053	-7,013	-6,937	-2,140	-1,853	-1,893	-2,200	-2,207	-2,170	-2,173

Fuente: Elaboración propia.

^a Se resaltan los valores de aquellos modelos que reportan el menor error medio.

Cuadro 26
Raíz del error cuadrático medio del pronóstico de pobreza en jóvenes^a

País	RMSE													
	Pobreza							Pobreza extrema						
	CEPAL	PVAR	PVAR+ exog	LASSO	LASSO+ exog	Pooling	Pooling	CEPAL	PVAR	PVAR+ exog	LASSO	LASSO+ exog	Pooling	Pooling
(0)	(1)	(2)	(3)	(5)	(0)+(3)	(0)+(5)	(0)	(1)	(2)	(3)	(5)	(0)+(3)	(0)+(5)	
Argentina	7,186	5,528	5,556	8,096	7,453	7,619	7,265	5,112	4,968	5,040	5,333	5,208	5,219	5,158
Bolivia (Estado Plurinacional de)	5,403	5,236	5,002	5,240	5,095	5,277	5,204	4,853	5,061	4,948	4,883	4,731	4,860	4,786
Brasil	11,551	8,307	9,217	11,889	11,804	11,715	11,671	4,389	3,539	3,756	4,462	4,532	4,424	4,458
Colombia	4,534	1,849	2,591	5,036	5,114	4,780	4,807	3,616	2,280	2,611	3,977	3,849	3,790	3,720
Costa Rica	8,444	6,694	7,291	8,952	8,729	8,686	8,575	3,029	2,711	2,852	3,196	3,234	3,110	3,129
Ecuador	4,367	3,265	3,640	5,093	4,980	4,706	4,639	1,935	1,798	1,721	1,473	1,708	1,583	1,679
El Salvador	7,937	5,519	6,352	8,632	8,793	8,282	8,359	3,581	2,507	2,943	3,956	3,863	3,765	3,718
Panamá	5,261	3,654	4,077	5,586	5,315	5,410	5,270	4,450	3,979	4,127	4,591	4,488	4,513	4,466
Paraguay	4,881	3,768	4,796	4,886	5,232	4,858	5,026	3,302	3,120	3,088	3,385	3,458	3,333	3,363
Perú	8,407	7,666	8,166	8,858	8,486	8,624	8,438	5,370	4,932	5,211	5,755	5,362	5,559	5,363
República Dominicana	5,880	5,754	6,110	6,394	6,484	6,112	6,145	3,325	2,979	3,010	3,442	3,627	3,362	3,446
Uruguay	8,654	7,344	7,558	9,080	8,999	8,863	8,819	3,794	3,416	3,508	3,852	3,902	3,822	3,847

Fuente: Elaboración propia.

^aSe resaltan los valores de aquellos modelos que reportan la menor raíz del error cuadrático medio.

Cuadro 27
Error medio absoluto del pronóstico de pobreza en jóvenes^a

País	Pobreza							Pobreza extrema						
	CEPAL	PVAR	PVAR+ exog	LASSO	LASSO+ exog	Pooling	Pooling	CEPAL	PVAR	PVAR+ exog	LASSO	LASSO+ exog	Pooling	Pooling
	(0)	(1)	(2)	(3)	(5)	(0)+(3)	(0)+(5)	(0)	(1)	(2)	(3)	(5)	(0)+(3)	(0)+(5)
Argentina	5,573	4,107	4,160	6,633	6,227	6,103	5,900	4,580	4,413	4,487	4,827	4,713	4,703	4,647
Bolivia (Estado Plurinacional de)	3,908	4,046	3,900	3,554	3,262	3,731	3,585	3,282	3,700	3,492	3,238	3,123	3,253	3,195
Brasil	11,007	7,733	8,587	11,260	11,313	11,133	11,160	4,253	3,387	3,587	4,327	4,393	4,290	4,323
Colombia	4,143	1,429	2,093	4,671	4,736	4,407	4,439	3,200	1,879	2,264	3,571	3,479	3,379	3,339
Costa Rica	7,888	5,953	6,500	8,200	8,006	8,044	7,947	2,747	2,359	2,506	2,894	2,935	2,821	2,841
Ecuador	3,200	2,721	2,743	3,929	3,836	3,536	3,489	1,600	1,436	1,450	1,257	1,379	1,400	1,461
El Salvador	7,638	4,838	5,777	8,346	8,508	7,992	8,073	3,015	2,092	2,431	3,369	3,338	3,177	3,162
Panamá	4,069	2,838	3,077	4,438	4,369	4,254	4,219	3,262	2,746	2,777	3,462	3,338	3,362	3,300
Paraguay	3,700	3,265	3,853	3,747	4,106	3,694	3,874	2,418	2,212	2,188	2,424	2,594	2,415	2,506
Perú	5,365	4,835	5,159	5,812	5,512	5,588	5,438	3,294	2,900	3,182	3,482	3,359	3,371	3,326
República Dominicana	4,659	4,971	5,235	4,976	4,912	4,806	4,744	2,759	2,441	2,441	2,882	2,953	2,821	2,856
Uruguay	6,820	5,407	5,640	7,207	7,053	7,013	6,937	2,193	1,907	1,947	2,213	2,220	2,203	2,207

Fuente: Elaboración propia.

^a Se resaltan los valores de aquellos modelos que reportan el menor error medio absoluto.

Cuadro 28
Error mediano absoluto del pronóstico de pobreza en jóvenes^a

País	Pobreza							Pobreza extrema						
	CEPAL	PVAR	PVAR+ exog	LASSO	LASSO+ exog	Pooling	Pooling	CEPAL	PVAR	PVAR+ exog	LASSO	LASSO+ exog	Pooling	Pooling
	(0)	(1)	(2)	(3)	(5)	(0)+(3)	(0)+(5)	(0)	(1)	(2)	(3)	(5)	(0)+(3)	(0)+(5)
Argentina	4,400	2,800	2,400	5,300	5,300	4,600	3,850	4,400	4,300	4,300	4,400	4,400	4,400	4,400
Bolivia (Estado Plurinacional de)	3,000	2,400	2,300	2,800	2,600	3,050	3,000	2,570	2,700	2,600	2,500	2,500	2,535	2,535
Brasil	12,200	7,800	8,500	12,100	12,000	12,350	12,150	3,900	3,200	3,400	4,100	4,000	3,950	3,850
Colombia	3,650	1,300	2,000	4,500	5,300	4,000	4,400	2,900	1,900	2,150	3,200	3,350	3,050	2,975
Costa Rica	8,200	6,100	6,200	8,200	8,200	8,200	8,200	2,800	2,600	2,600	2,800	3,000	2,800	2,850
Ecuador	1,900	2,400	2,150	2,550	3,200	1,950	2,250	1,500	1,150	1,300	1,300	1,200	1,675	1,675
El Salvador	7,400	3,900	5,100	8,000	8,100	7,700	7,750	2,900	1,600	2,400	3,000	3,000	2,900	2,900
Panamá	2,100	2,000	1,600	2,200	2,700	2,150	2,900	1,800	1,600	1,000	1,900	1,800	1,800	1,800
Paraguay	2,700	3,400	3,500	2,700	3,800	2,700	2,850	1,500	1,400	1,300	1,400	1,500	1,400	1,500
Perú	2,600	2,200	2,100	3,000	3,000	2,800	2,450	2,200	1,700	1,300	2,400	2,200	2,300	2,200
República Dominicana	3,300	4,100	6,100	4,900	2,500	3,950	2,550	2,100	2,200	2,000	2,300	3,000	2,150	2,650
Uruguay	4,600	3,100	3,600	4,900	5,200	4,750	4,900	0,500	0,300	0,300	0,500	0,500	0,500	0,500

Fuente: Elaboración propia.

^aSe resaltan los valores de aquellos modelos que reportan el menor error mediano absoluto.

VII. Recomendaciones para la práctica

Este trabajo explota al máximo los esfuerzos para mejorar todas las etapas involucradas en el ejercicio predictivo, pero respetando los insumos de información usados en la estrategia habitualmente implementada por CEPAL. Es decir, se focaliza en las potenciales mejoras *metodológicas* aplicables al mismo conjunto de información. Naturalmente, la posibilidad de aumentar el conjunto de información o buscar fuentes adicionales podría redundar en mejoras en la capacidad predictiva.

En el marco de la metodología “micro-macro” adoptada por CEPAL, un insumo clave es disponer de buenos pronósticos del PIB y de los cambios en el coeficiente de Gini. Aquí aparecen varios resultados relevantes. Las distintas variantes de los modelos implementados explotan: a) la variabilidad temporal de las variables involucradas, b) la “complejidad” de los modelos. En relación con el primer punto, varios de los modelos implementados explotan el hecho de que tanto el PIB como el coeficiente de Gini parecen tener algún vínculo intertemporal con sus valores pasados. Como la información dentro de cada país es limitada (solo unos pocos años), una estrategia intermedia entre estimar modelos por países y un solo modelo temporal para todos los países es implementar un Panel VAR, en el que la incorporación de efectos fijos permite acomodar heterogeneidades a nivel de país y la incorporación de unos pocos rezagos permite explotar la naturaleza autorregresiva de las variables. Aun cuando parsimoniosa, la estructura VAR sumada a la de efectos fijos rápidamente agota los grados de libertad disponibles. A tal fin, se implementaron técnicas de *machine learning* (como LASSO o Elastic Net) para “regularizar” el Panel VAR y así ahorrar grados de libertad en términos de la inclusión de efectos fijos y rezagos.

Estos ejercicios sugieren varios resultados interesantes. En primer lugar, y aun cuando la dimensión temporal por país es corta, el pasado del PIB y del Gini ayudan a su predicción. Es decir, en los insumos disponibles hay información “temporal” que permite mejorar el pronóstico de estas variables, lo cual es esperable, teniendo en cuenta la naturaleza relativamente suave de la evolución de estas variables, con un fuerte componente tendencial y, en segundo lugar, cíclico. Esta es precisamente la idea que intenta captar el modelo “PVAR regularizado”. Efectivamente, se observa una mejora en la capacidad predictiva del crecimiento del PIB y la variación del Gini en relación a los insumos utilizados

por CEPAL. No obstante, estas mejoras en la predicción de los “insumos” no se traducen en diferencias cuantitativas apreciables en la proyección de la pobreza. Una explicación posible a este hecho es que la fórmula de actualización usada por CEPAL (la que proyecta ingresos de individuos en base a información micro y macro) tiende a explotar mayoritariamente el cambio en el PIB por sobre lo que sucede con el coeficiente de Gini. Asimismo, en las exploraciones implementadas en este trabajo (en particular, en las que no restringen ex-ante la relación entre la pobreza y sus determinantes (el PIB y el coeficiente de Gini), la importancia relativa del coeficiente de Gini en la predicción de la pobreza es baja. O sea que las ganancias de pronóstico en ambas variables solo se manifiestan en relación a la eventual mejora en la predicción del PIB.

El trabajo presenta una segunda alternativa posible que consiste en abandonar el uso de microdatos de las encuestas (enfoque “micro macro”) y saltar directamente a una proyección “macro”, es decir, intentar proyectar la tasa de pobreza de un país directamente a partir de los pronósticos de PIB y Gini. Los resultados muestran que el enfoque puramente “macro” genera predicciones muy similares a las del enfoque “micro macro”, de lo que se deduce que la etapa “micro” es fundamentalmente una especie de “regla de cómputo” para el cálculo de la predicción de la pobreza. Este resultado es importante para eventuales reformulaciones de las metodologías de predicción, ya que dice que, en su operatoria actual, es poco lo que agrega la etapa “micro” y potencialmente mucho lo que podría agregar incorporar información agregada que mejore la capacidad predictiva.

En esta línea, se intentó agregar información puramente macro para mejorar la capacidad predictiva, como la relacionada con el tipo de cambio global. La mejora predictiva es leve, lo cual, a la luz de la discusión anterior, no descarta que información macro a nivel de país no redunde en mejora sustanciales. La posibilidad de agregar información a nivel país posiblemente permita captar mejor la naturaleza cíclica y episódica de la pobreza. Naturalmente, esta posibilidad requiere un esfuerzo de relevar y compaginar dicha información. El tipo de variables que posiblemente contribuya a mejorar las predicciones debería reaccionar a la naturaleza cíclica de la economía debería, refinar lo que en este sentido capta el PBI. Muy posiblemente variables específicas del mercado laboral (desempleo, salarios, etc.) o de actividad sectorial puedan ayudar en este sentido. Misma cosa con factores fijos, que en este trabajo fueron agrupados a nivel de país. Posiblemente datos de las características específicas de estos grupos (como su nivel de educación o algunas características de las viviendas o de los servicios de salud) puedan contribuir a refinar el trabajo que en este informe han cumplido los efectos fijos.

En relación con la capacidad predictiva, se implementaron varios ejercicios basados en la idea de crear “pseudo-individuos”, es decir, ante la imposibilidad de observar a una misma persona en varios periodos, se construyeron grupos de individuos (ej. mujeres, desocupados y jóvenes), naturalmente observables en varios períodos y que permiten evaluar el desempeño predictivo de los modelos, con los resultados antes comentados. Naturalmente, la posibilidad de usar información de paneles de micro datos (no disponibles para esta investigación) es una importante ruta a explorar.

En términos generales, la exhaustiva evaluación realizada en este trabajo muestra que los modelos flexibles, como los que surgen de LASSO y PVAR prometen mejoras cuantitativas, al explotar formas funcionales alternativas a la usada por CEPAL y también la dinámica de los predictores. Así y todo, cuantitativamente los resultados implementados por la metodología de CEPAL a la fecha, sin bien cualitativamente inferiores, son un competidor razonable.

Respecto a la posibilidad de utilizar modelos de predicción más complejos, cabe tener presente que el uso de métodos más flexibles, propios del enfoque de “machine learning” se ve limitado por la limitada cantidad de información disponible, dada la frecuencia y la cantidad de países para los cuales se hace el pronóstico. Más concretamente, para el tipo de información disponible (la cantidad de predictores) aumentar la complejidad de los modelos es una tarea delicada ya que se cae rápidamente en el problema de sobreajustar (*overfit*) los datos disponibles a costa de la calidad de predicción futura.

En los ejercicios implementados, el uso de métodos flexibles, como LASSO y Elastic Net, requirió cierto esfuerzo para evitar esta dificultad, y las mejoras fueron cualitativamente relevantes si bien cuantitativamente menores. En síntesis, para el conjunto de información disponible y para la frecuencia y cantidad de países a pronosticar, hay poco espacio para explorar la ruta de la complejidad de los modelos predictivos.

Otra alternativa es explorar la estructura de *panel* implícita en algunas encuestas de hogares, o a nivel de pseudo individuos. Concretamente, la literatura de “persistencia de ingresos”, por ejemplo, Arias et al. (2011) explota la dinámica de ingresos individuales y encuentra que su dinámica tiene un fuerte componente de persistencia asociado a su propio pasado y a las características invariantes de los individuos. Es decir, y en relación a lo implementado en este trabajo, se trata de implementar métodos parecidos a los de paneles dinámicos, pero a nivel individual o de hogar. Este tipo de enfoque (verdaderamente “micro”) permitiría predecir tanto tasas agregadas como individuales o de grupos específicos. Una vez más, la limitación obvia de esta aproximación se relaciona con la dificultad (cuando no imposibilidad) de construir auténticos paneles a nivel micro.

Finalmente, una ruta adicional consiste en explotar información “no estándar” sobre el bienestar como la basada en indicadores online como en Chetty et al. (2020). Este tipo de aproximación explota información online, fundamentalmente del sector privado, para construir indicadores que permiten monitorear, en tiempo real, el funcionamiento de la economía. Es importante aclarar que este tipo de ejercicios no pretende predecir la pobreza o el bienestar medidos en sus formas “estándar” sino construir otros indicadores que faciliten el monitoreo, con una frecuencia más alta o con menor rezago que el que naturalmente tienen las oficinas públicas de estadísticas. Es decir, no necesariamente se debería esperar una mejora en el pronóstico de la pobreza de seguir esta ruta, sino proveer una forma alternativa de medirla, tal vez con mayor frecuencia y especificidad. Naturalmente, la implementación de estas ideas requiere un gran esfuerzo de compatibilización, complejo de implementar aun para un solo país. Estas complejidades aumentan cuando se trata de países en desarrollo, como los estudiados en este trabajo. También cabe aclarar que, a diferencia de las cuestiones financieras o macroeconómicas, el aumento en la frecuencia de las mediciones no es una característica necesariamente deseable para la pobreza o la desigualdad, toda vez que estas tienden a moverse en forma suave, por lo que una medición a una frecuencia mayor a la deseable posiblemente venga acompañada de una considerable dosis de “ruido” estadístico.

Comentarios finales

La batería de ejercicios predictivos implementados en este trabajo sugiere algunas conclusiones relevantes para el problema del pronóstico de la pobreza en América Latina.

Es importante resaltar el punto de partida y las restricciones del problema enfrentado. El objetivo genérico consiste en predecir la tasa de pobreza agregada (y la tasa de pobreza extrema) para un conjunto de países, en diversos períodos. La información de base disponible consiste en: 1) microdatos de ingresos de hogares, para varios países y períodos, 2) información agregada de los niveles y cambios en el ingreso medio y en el coeficiente de Gini. El modelo base 1) toma los ingresos de los microdatos, 2) realiza un pronóstico de estos, en base a cambios en proyecciones del ingreso medio y del coeficiente Gini, y 3) la tasa de pobreza pronosticada es la tasa de pobreza computada en base a estos ingresos predichos, en relación con una línea de pobreza también proyectada.

En primer lugar, cabe reflexionar acerca del conjunto de información disponible para esta tarea. Como adelantásemos, se trata de un ejercicio "micro-macro". "Macro" porque tanto el resultado del modelo (la tasa de pobreza) como dos de sus insumos clave (el PIB y el Gini) se mueven a ese nivel de agregación (por país). La parte "micro" del problema tiene que ver con que la actualización del pronóstico se basa en pronósticos de micro datos, que luego son agregados nuevamente para computar una tasa de pobreza.

En segundo lugar, un punto recurrente en este trabajo consiste en fijar algún punto de referencia (*benchmark*) que permita evaluar el desempeño predictivo de los modelos utilizados. Como comentásemos anteriormente, la tarea no es simple, toda vez que los pronósticos de ingreso se hacen a nivel "micro", pero el mismo individuo no es observable en varias encuestas, de modo que resulta imposible realizar una evaluación de la capacidad predictiva a ese nivel. Este trabajo implementa varios ejercicios, a distintos niveles de agregación, que permiten tener una idea aproximada de la capacidad predictiva de los modelos implementados.

La naturaleza relativamente tendencial y levemente cíclica de la pobreza hace que pueda caracterizarse razonablemente con datos de similar nivel de agregación y frecuencia, como el PIB o el coeficiente de Gini. Los resultados de este trabajo sugieren que hay espacio para las mejoras predictivas si además se incorporan aspectos dinámicos, propios de proceso estocásticos con una obvia “memoria” como las series antes mencionadas. Asimismo, la incorporación de información adicional de naturaleza “temporal” (como la del mercado cambiario) puede asistir a la mejora predictiva, máxime a nivel de cada país, como se sugiriese en el trabajo, y que queda como una ruta de investigación futura relevante.

Fuera del marco de este trabajo queda la posibilidad de predecir la condición de pobreza a nivel *individual* y juzgar el desempeño predictivo a ese nivel. La disponibilidad de microdatos a través de encuestas de hogares parecen ser un terreno fértil para este tipo de ejercicios, como se reporta en el reciente estudio de Sosa Escudero et al. (2021), pero, por las razones argumentadas en este trabajo, no es obvio que estos ejercicios impliquen una mejora a fines de predecir la pobreza a nivel *agregado* —de país en un periodo específico— como es el objetivo de este trabajo.

Bibliografía

- Aguilar, R. A. C., Mahler, D. G., & Newhouse, D. (2019), "Nowcasting Global Poverty", *Special IARIW-World Bank Conference New Approaches to Defining and Measuring Poverty in a Growing World*, Washington, DC, November 7-8.
- Aiken, E. L., Bedoya, G., Coville, A., & Blumenstock, J. E. (2020), *Targeting Development Aid with Machine Learning and Mobile Phone Data*.
- Arias, O., Marchionni, M., & Sosa Escudero, W. (2011), "Sources of Income Persistence: Evidence from Rural El Salvador", 2011, *Journal of Income Distribution*, vol. 20(1), 3-28.
- Blumenstock, J. E. (2018), "Estimating economic characteristics with phone data". En *AEA papers and proceedings* (Vol. 108, pp. 72-76).
- Blumenstock, J., Cadamuro, G., & On, R. (2015), "Predicting poverty and wealth from mobile phone metadata". *Science*, 350(6264), 1073-1076.
- Brum, M., & De Rosa, M. (2021), "Too little but not too late: nowcasting poverty and cash transfers' incidence during COVID-19's crisis". *World Development*, 140, 105227.
- Caruso, G., Sosa-Escudero, W., & Svarc, M. (2015), "Deprivation and the dimensionality of welfare: a variable-selection cluster-analysis approach". *Review of Income and Wealth*, 61(4), 702-722.
- CEPAL, IPEA & PNUD. (2003), *Hacia el objetivo del milenio de reducir la pobreza en América Latina y el Caribe*, Santiago de Chile, febrero.
- Chetty, R., Friedman, J., Hendren, N. & Stepner, M. (2020), *The Economic Impacts of COVID-19: Evidence from a New Public Database Built Using Private Sector Data*, working paper.
- Edo, M., Escudero, W. S., & Svarc, M. (2021), "A multidimensional approach to measuring the middle class". *The Journal of Economic Inequality*, 19(1), 139-162.
- Elbers, C., Lanjouw, J. O., & Lanjouw, P. (2003), "Micro-level estimation of poverty and inequality". *Econometrica*, 71(1), 355-364.
- Elliott, G., & Timmermann, A. (2016), "Economic Forecasting". *New Jersey: Princeton University Press*.
- Giannone, D., Reichlin, L., & Small, D. (2008), "Nowcasting: The real-time informational content of macroeconomic data". *Journal of Monetary Economics*, 55(4), 665-676.
- Hendry, D., & Clements, M. (2004), "Pooling of forecasts". *The Econometrics Journal*, 7(1), 1-31.

- Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016), "Combining satellite imagery and machine learning to predict poverty". *Science*, 353(6301), 790-794.
- Kleinberg, J., Ludwig, J., Mullainathan, S. & Obermeyer, Z. (2015), "Prediction policy problems". *American Economic Review*, 105(5), 491-495.
- Lucchetti, L. (2018), *What can we (machine) learn about welfare dynamics from cross-sectional data?* The World Bank.
- Sampi Bravo, J. R. E., & Jooste, C. (2020), "Nowcasting economic activity in times of COVID-19: An approximation from the Google Community Mobility Report". *World Bank Policy Research Working Paper*, (9247).
- Sosa Escudero, W. (2019), "Big data y aprendizaje automático: Ideas y desafíos para economistas", Capítulo 5 en Ahumada, Gabrielli, Herrera Gomez y Sosa Escudero, 2019, *Una Nueva Econometría: Automatización, Big data, Econometría Estructural y Espacial*, Editorial de la Universidad Nacional del Sur.
- Sosa Escudero, W., Anauati, V. & Brau, W. (2021), "Poverty, inequality and development studies with machine learning", en Matyas, L. y Chan, F. (eds) *Econometrics with Machine Learning*, Springer, New York (en prensa).
- Tibshirani, R. (1996), "Regression shrinkage and selection via lasso". *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288.

Anexos

Anexo 1

Matrices de confusión por país para el año 2019

Cuadro A1
Argentina

Agregación en media

observado	proyectado		Total
	0	1	
0	55	5	60
1	5	10	15
Total	60	15	75

Sensitivity	D/(C+D)	66.67%
Specificity	A/(A+B)	91.67%
Positive predictive value	D/(B+D)	66.67%
Negative predictive value	A/(A+C)	91.67%
False positive rate	B/(A+B)	8.33%
False negative rate	C/(C+D)	33.33%
Correctly classified	A+D/(A+B+C+D)	86.67%
Effect strength for sensitivity		58.33%
ROC area		0.7917

Agregación en mediana

observado	proyectado		Total
	0	1	
0	38	9	47
1	11	17	28
Total	49	26	75

Sensitivity	D/(C+D)	60.71%
Specificity	A/(A+B)	80.85%
Positive predictive value	D/(B+D)	65.38%
Negative predictive value	A/(A+C)	77.55%
False positive rate	B/(A+B)	19.15%
False negative rate	C/(C+D)	39.29%
Correctly classified	A+D/(A+B+C+D)	73.33%
Effect strength for sensitivity		41.57%
ROC area		0.7078

Fuente: Elaboración propia.

Cuadro A2
Bolivia

Agregación en media

observado	proyectado		Total
	0	1	
0	41	3	44
1	1	5	6
Total	42	8	50

Sensitivity	D/(C+D)	83.33%
Specificity	A/(A+B)	93.18%
Positive predictive value	D/(B+D)	62.50%
Negative predictive value	A/(A+C)	97.62%
False positive rate	B/(A+B)	6.82%
False negative rate	C/(C+D)	16.67%
Correctly classified	A+D/(A+B+C+D)	92.00%
Effect strength for sensitivity		76.52%
ROC area		0.8826

Agregación en mediana

observado	proyectado		Total
	0	1	
0	28	5	33
1	1	16	17
Total	29	21	50

Sensitivity	D/(C+D)	94.12%
Specificity	A/(A+B)	84.85%
Positive predictive value	D/(B+D)	76.19%
Negative predictive value	A/(A+C)	96.55%
False positive rate	B/(A+B)	15.15%
False negative rate	C/(C+D)	5.88%
Correctly classified	A+D/(A+B+C+D)	88.00%
Effect strength for sensitivity		78.97%
ROC area		0.8948

Fuente: Elaboración propia.

Cuadro A3
Brasil

Agregación en media

observado	proyectado		Total
	0	1	
0	55	1	56
1	1	2	3
Total	56	3	59

Sensitivity	D/(C+D)	66.67%
Specificity	A/(A+B)	98.21%
Positive predictive value	D/(B+D)	66.67%
Negative predictive value	A/(A+C)	98.21%
False positive rate	B/(A+B)	1.79%
False negative rate	C/(C+D)	33.33%
Correctly classified	A+D/(A+B+C+D)	96.61%
Effect strength for sensitivity		64.88%
ROC area		0.8244

Fuente: Elaboración propia.

Agregación en mediana

observado	proyectado		Total
	0	1	
0	51	1	52
1	1	6	7
Total	52	7	59

Sensitivity	D/(C+D)	85.71%
Specificity	A/(A+B)	98.08%
Positive predictive value	D/(B+D)	85.71%
Negative predictive value	A/(A+C)	98.08%
False positive rate	B/(A+B)	1.92%
False negative rate	C/(C+D)	14.29%
Correctly classified	A+D/(A+B+C+D)	96.61%
Effect strength for sensitivity		83.79%
ROC area		0.9190

Cuadro A4
Costa Rica

Agregación en media

observado	proyectado		Total
	0	1	
0	52	1	53
1	2	0	2
Total	54	1	55

Sensitivity	D/(C+D)	0.00%
Specificity	A/(A+B)	98.11%
Positive predictive value	D/(B+D)	0.00%
Negative predictive value	A/(A+C)	96.30%
False positive rate	B/(A+B)	1.89%
False negative rate	C/(C+D)	100.00%
Correctly classified	A+D/(A+B+C+D)	94.55%
Effect strength for sensitivity		-1.89%
ROC area		0.4906

Fuente: Elaboración propia.

Agregación en mediana

observado	proyectado		Total
	0	1	
0	46	1	47
1	6	2	8
Total	52	3	55

Sensitivity	D/(C+D)	25.00%
Specificity	A/(A+B)	97.87%
Positive predictive value	D/(B+D)	66.67%
Negative predictive value	A/(A+C)	88.46%
False positive rate	B/(A+B)	2.13%
False negative rate	C/(C+D)	75.00%
Correctly classified	A+D/(A+B+C+D)	87.27%
Effect strength for sensitivity		22.87%
ROC area		0.6144

Cuadro A5
Ecuador

Agregación en media

observado	proyectado		Total
	0	1	
0	39	0	39
1	6	2	8
Total	45	2	47

Sensitivity	D/(C+D)	25.00%
Specificity	A/(A+B)	100.00%
Positive predictive value	D/(B+D)	100.00%
Negative predictive value	A/(A+C)	86.67%
False positive rate	B/(A+B)	0.00%
False negative rate	C/(C+D)	75.00%
Correctly classified	A+D/(A+B+C+D)	87.23%
Effect strength for sensitivity		25.00%
ROC area		0.6250

Agregación en mediana

observado	proyectado		Total
	0	1	
0	36	1	37
1	2	8	10
Total	38	9	47

Sensitivity	D/(C+D)	80.00%
Specificity	A/(A+B)	97.30%
Positive predictive value	D/(B+D)	88.89%
Negative predictive value	A/(A+C)	94.74%
False positive rate	B/(A+B)	2.70%
False negative rate	C/(C+D)	20.00%
Correctly classified	A+D/(A+B+C+D)	93.62%
Effect strength for sensitivity		77.30%
ROC area		0.8865

Fuente: Elaboración propia.

Cuadro A6
El Salvador

Agregación en media

observado	proyectado		Total
	0	1	
0	41	4	45
1	1	5	6
Total	42	9	51

Sensitivity	D/(C+D)	83.33%
Specificity	A/(A+B)	91.11%
Positive predictive value	D/(B+D)	55.56%
Negative predictive value	A/(A+C)	97.62%
False positive rate	B/(A+B)	8.89%
False negative rate	C/(C+D)	16.67%
Correctly classified	A+D/(A+B+C+D)	90.20%
Effect strength for sensitivity		74.44%
ROC area		0.8722

Agregación en mediana

observado	proyectado		Total
	0	1	
0	31	6	37
1	1	13	14
Total	32	19	51

Sensitivity	D/(C+D)	92.86%
Specificity	A/(A+B)	83.78%
Positive predictive value	D/(B+D)	68.42%
Negative predictive value	A/(A+C)	96.88%
False positive rate	B/(A+B)	16.22%
False negative rate	C/(C+D)	7.14%
Correctly classified	A+D/(A+B+C+D)	86.27%
Effect strength for sensitivity		76.64%
ROC area		0.8832

Fuente: Elaboración propia.

**Cuadro A7
Honduras****Agregación en media**

observado	proyectado		Total
	0	1	
0	32	3	35
1	7	19	26
Total	39	22	61

Sensitivity	D/(C+D)	73.08%
Specificity	A/(A+B)	91.43%
Positive predictive value	D/(B+D)	86.36%
Negative predictive value	A/(A+C)	82.05%
False positive rate	B/(A+B)	8.57%
False negative rate	C/(C+D)	26.92%
Correctly classified	A+D/(A+B+C+D)	83.61%
Effect strength for sensitivity		64.51%
ROC area		0.8225

Agregación en mediana

observado	proyectado		Total
	0	1	
0	24	4	28
1	4	29	33
Total	28	33	61

Sensitivity	D/(C+D)	87.88%
Specificity	A/(A+B)	85.71%
Positive predictive value	D/(B+D)	87.88%
Negative predictive value	A/(A+C)	85.71%
False positive rate	B/(A+B)	14.29%
False negative rate	C/(C+D)	12.12%
Correctly classified	A+D/(A+B+C+D)	86.89%
Effect strength for sensitivity		73.59%
ROC area		0.8680

Fuente: Elaboración propia.

**Cuadro A8
Panamá****Agregación en media**

observado	proyectado		Total
	0	1	
0	54	2	56
1	2	1	3
Total	56	3	59

Sensitivity	D/(C+D)	33.33%
Specificity	A/(A+B)	96.43%
Positive predictive value	D/(B+D)	33.33%
Negative predictive value	A/(A+C)	96.43%
False positive rate	B/(A+B)	3.57%
False negative rate	C/(C+D)	66.67%
Correctly classified	A+D/(A+B+C+D)	93.22%
Effect strength for sensitivity		29.76%
ROC area		0.6488

Agregación en mediana

observado	proyectado		Total
	0	1	
0	52	2	54
1	2	3	5
Total	54	5	59

Sensitivity	D/(C+D)	60.00%
Specificity	A/(A+B)	96.30%
Positive predictive value	D/(B+D)	60.00%
Negative predictive value	A/(A+C)	96.30%
False positive rate	B/(A+B)	3.70%
False negative rate	C/(C+D)	40.00%
Correctly classified	A+D/(A+B+C+D)	93.22%
Effect strength for sensitivity		56.30%
ROC area		0.7815

Fuente: Elaboración propia.

Cuadro A9
Paraguay

Agregación en media

observado	proyectado		Total
	0	1	
0	46	1	47
1	2	1	3
Total	48	2	50

Sensitivity	D/(C+D)	33.33%
Specificity	A/(A+B)	97.87%
Positive predictive value	D/(B+D)	50.00%
Negative predictive value	A/(A+C)	95.83%
False positive rate	B/(A+B)	2.13%
False negative rate	C/(C+D)	66.67%
Correctly classified	A+D/(A+B+C+D)	94.00%
Effect strength for sensitivity		31.21%
ROC area		0.6560

Agregación en mediana

observado	proyectado		Total
	0	1	
0	42	2	44
1	3	3	6
Total	45	5	50

Sensitivity	D/(C+D)	50.00%
Specificity	A/(A+B)	95.45%
Positive predictive value	D/(B+D)	60.00%
Negative predictive value	A/(A+C)	93.33%
False positive rate	B/(A+B)	4.55%
False negative rate	C/(C+D)	50.00%
Correctly classified	A+D/(A+B+C+D)	90.00%
Effect strength for sensitivity		45.45%
ROC area		0.7273

Fuente: Elaboración propia.

Cuadro A10
Perú

Agregación en media

observado	proyectado		Total
	0	1	
0	80	2	82
1	4	0	4
Total	84	2	86

Sensitivity	D/(C+D)	0.00%
Specificity	A/(A+B)	97.56%
Positive predictive value	D/(B+D)	0.00%
Negative predictive value	A/(A+C)	95.24%
False positive rate	B/(A+B)	2.44%
False negative rate	C/(C+D)	100.00%
Correctly classified	A+D/(A+B+C+D)	93.02%
Effect strength for sensitivity		-2.44%
ROC area		0.4878

Agregación en mediana

observado	proyectado		Total
	0	1	
0	72	8	80
1	3	3	6
Total	75	11	86

Sensitivity	D/(C+D)	50.00%
Specificity	A/(A+B)	90.00%
Positive predictive value	D/(B+D)	27.27%
Negative predictive value	A/(A+C)	96.00%
False positive rate	B/(A+B)	10.00%
False negative rate	C/(C+D)	50.00%
Correctly classified	A+D/(A+B+C+D)	87.21%
Effect strength for sensitivity		40.00%
ROC area		0.7000

Fuente: Elaboración propia.

Cuadro A11
República Dominicana

Agregación en media

Agregación en mediana

(no se identifican grupos pobres)

observado	proyectado		Total
	0	1	
0	55	3	58
1	2	2	4
Total	57	5	62

Sensitivity	D/(C+D)	50.00%
Specificity	A/(A+B)	94.83%
Positive predictive value	D/(B+D)	40.00%
Negative predictive value	A/(A+C)	96.49%
False positive rate	B/(A+B)	5.17%
False negative rate	C/(C+D)	50.00%
Correctly classified	A+D/(A+B+C+D)	91.94%
Effect strength for sensitivity		44.83%
ROC area		0.7241

Fuente: Elaboración propia.

Cuadro A12
Uruguay

Agregación en media

Agregación en mediana

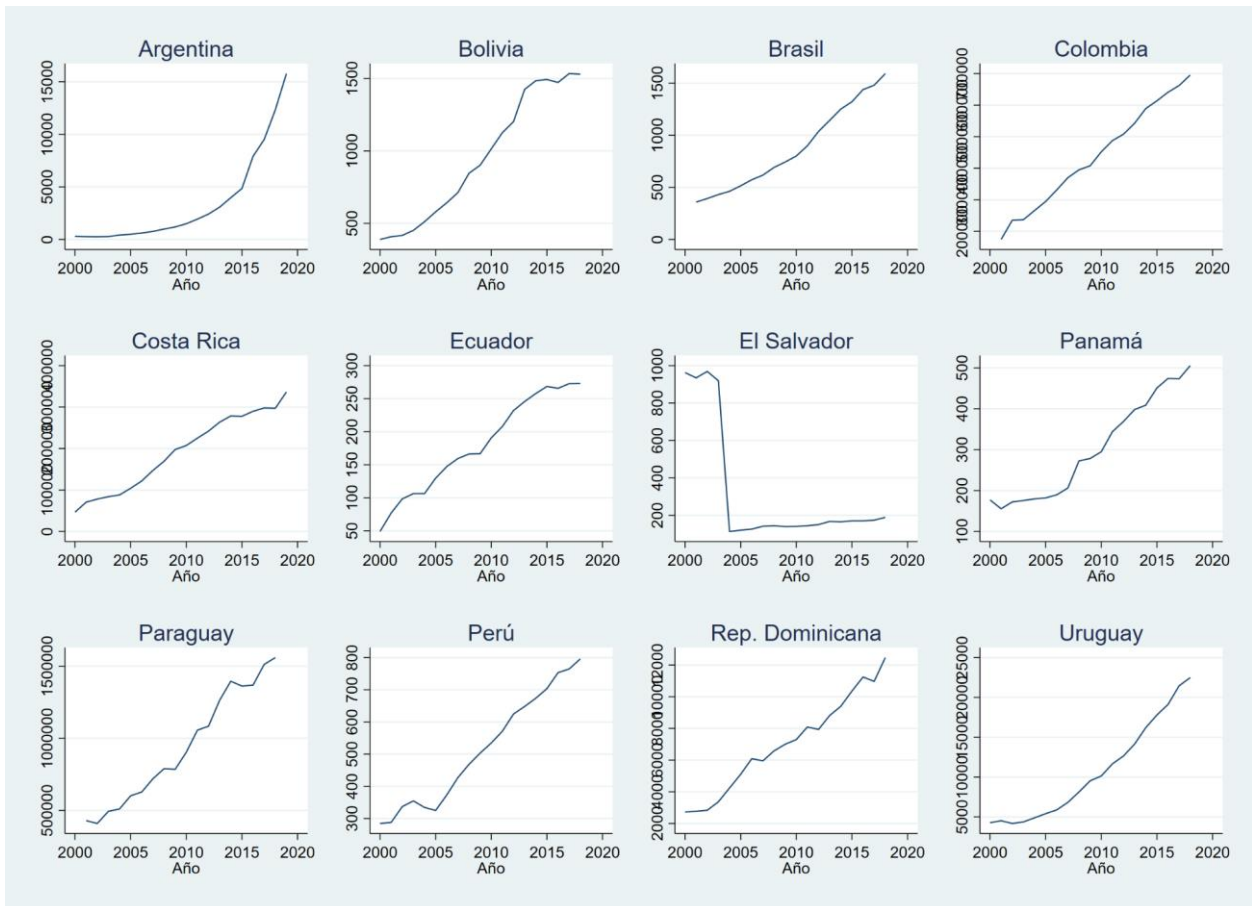
(no se identifican grupos pobres)

(no se identifican grupos pobres)

Fuente: Elaboración propia.

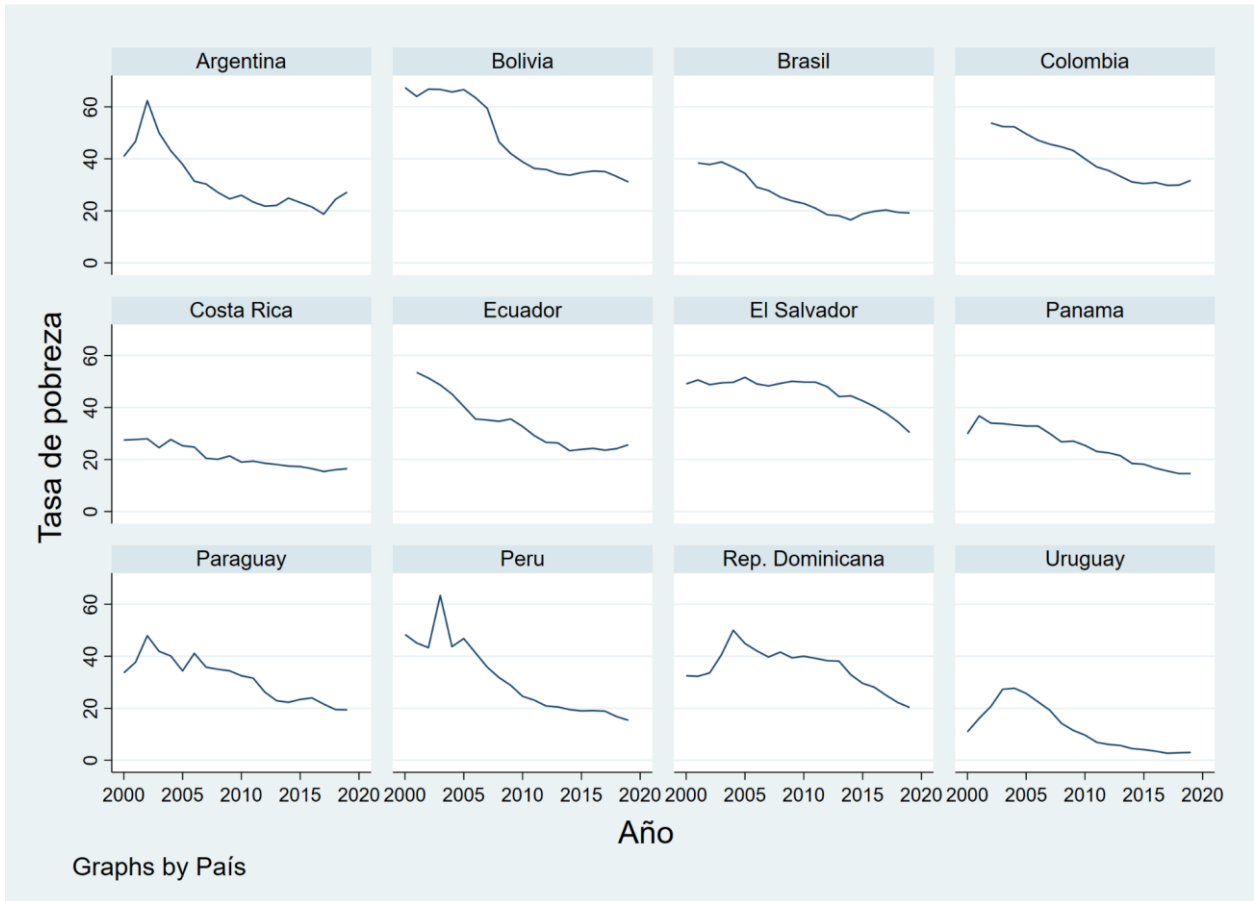
Anexo 2 Evolución de temporal de las variables relevadas

Gráfico A1
Evolución del ingreso medio per cápita (en moneda nacional)



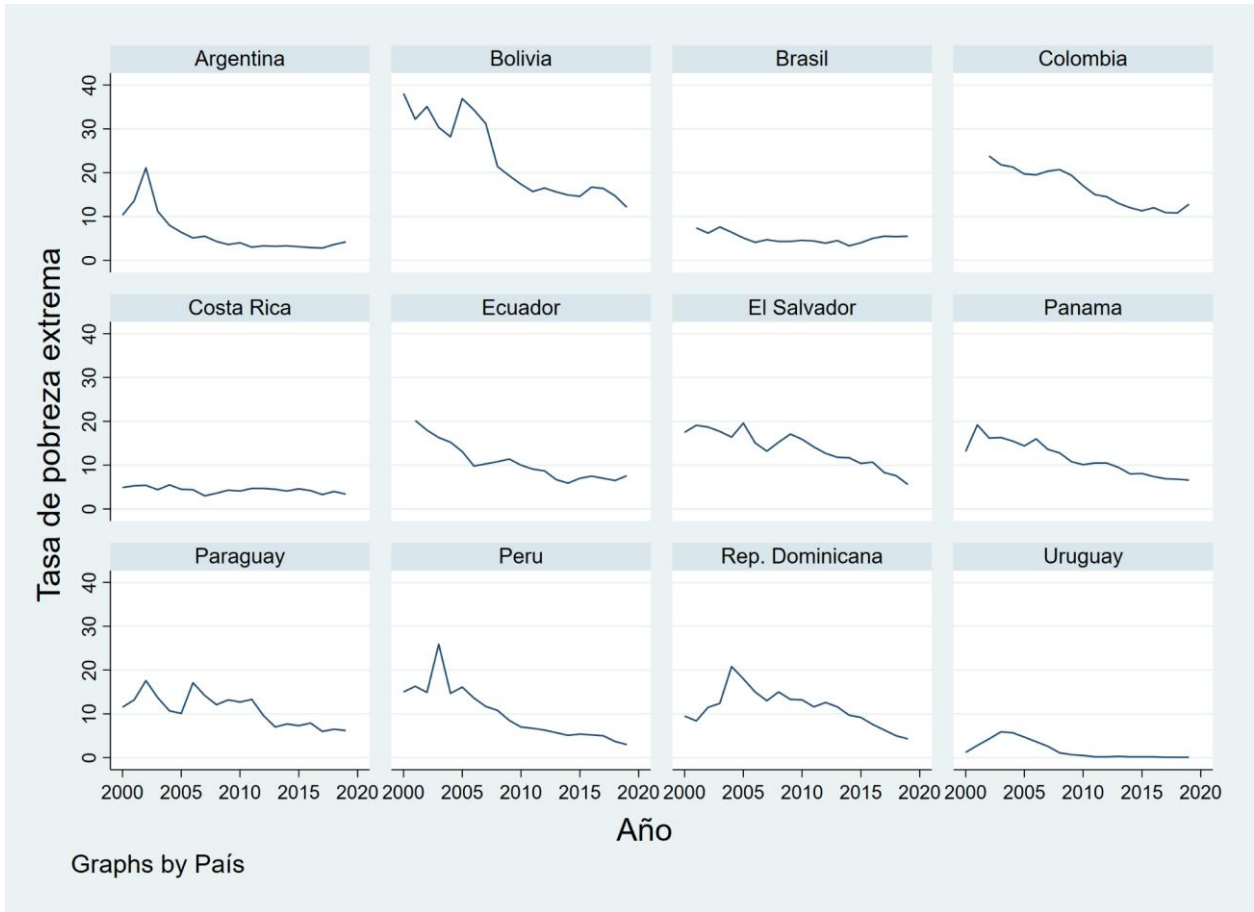
Fuente: Elaboración propia.

Gráfico A2
Evolución de la tasa de pobreza



Fuente: Elaboración propia.

Gráfico A3
Evolución de la tasa de pobreza extrema



Fuente: Elaboración propia.



NACIONES UNIDAS

Serie

CEPAL

Estudios Estadísticos

Números publicados

Un listado completo, así como los archivos pdf están disponibles en
www.cepal.org/publicaciones

103. Predicciones agregadas de pobreza con información a escala micro y macro: evaluación, diagnóstico y propuestas, Walter Sosa Escudero y Magdalena Cornejo, (LC/TS.2022/95), 2022.
102. La medición de la discriminación en base al autorreporte: estado de situación y desafíos, Pablo Villatoro, (LC/TS.2021/87), 2021.
101. Criterios de calidad en la estimación de indicadores a partir de encuestas de hogares: una aplicación a la migración internacional, Andrés Gutierrez, Xavier Mancero, Alvaro Fuentes, Felipe López, Felipe Molina, (LC/TS.2020/52), 2020.
100. Desafíos en el diseño de medidas de pobreza multidimensional, María Emma Santos, (LC/TS.2019/5), 2019.
99. Non monetary indicators to monitor SDG targets 1.2 and 1.4: Standards, availability, comparability and quality, María Emma Santos, (LC/TS.2019/4), 2018.
98. Una propuesta de estimación del producto interno bruto trimestral de América Latina y el Caribe, Giannina López, Alberto Malmierca, (LC.TS.2018/88), 2018.
97. Desagregación de datos en encuestas de hogares. Metodologías de estimación en áreas pequeñas, Isabel Molina, (LC/TS.2018/82), 2018.
96. ¿Cuál es el alcance de las transferencias no contributivas en América Latina?: discrepancias entre encuestas y registros, Pablo Villatoro, Simone Cecchini, (LC/TS.2018/46), 2018.
95. Avances y desafíos de las cuentas económico-ambientales en América Latina y el Caribe, Franco Carvajal, (LC/TS.2017/148), 2018.
94. La situación de las estadísticas, indicadores y cuentas ambientales en América Latina y el Caribe, (LC/TS.2017/135), 2017.

ESTUDIOS ESTADÍSTICOS

Números publicados:

- 103 Predicciones agregadas de pobreza con información a escala micro y macro
Evaluación, diagnóstico y propuestas
Walter Sosa Escudero y Magdalena Cornejo
- 102 La medición de la discriminación en base al autorreporte
Estado de situación y desafíos
Pablo Villatoro
- 101 Criterios de calidad en la estimación de indicadores a partir de encuestas de hogares
Una aplicación a la migración internacional
Andrés Gutiérrez, Xavier Mancero, Alvaro Fuentes, Felipe López y Felipe Molina
- 100 Desafíos en el diseño de medidas de pobreza multidimensional
María Emma Santos

