

Distribución del ingreso

Una aproximación geográfica al fenómeno del ingreso

Daniel Franzani C. · Álvaro Humeres
B. · Álvaro Paredes L. · Pablo Torres
I. ·

Received: date / Accepted: date

Abstract En este trabajo se busca generar un proxy entre un modelo lineal para explicar el ingreso en base a la CASEN 2017, utilizar dicho modelo con las variables homologadas del censo 2017 y georreferenciar los resultados, de manera de agregar mayor granularidad la información.

Keywords CASEN · CENSO · ingreso ·

1 Introducción

Durante el último tiempo, Chile se ha enfrentado a un amplio debate acerca de los altos indicadores de desigualdad que posee el país (ver [1] y [2]).

Una de las principales manifestaciones de este ámbito es la zona urbana, donde la ciudad de Santiago se ha visto caracterizada como una metrópolis de fuertes diferencias en la calidad de vida, en las diferentes comunas que la conforman [3].

Daniel Franzani C.
Pontificia Universidad Católica de Chile
E-mail: dsfranza@uc.cl

Álvaro Humeres B.
Pontificia Universidad Católica de Chile
E-mail: adhumeres@uc.cl

Álvaro Paredes L.
Pontificia Universidad Católica de Chile
E-mail: aaparede@uc.cl

Pablo Torres I.
Pontificia Universidad Católica de Chile
E-mail: patorre4@uc.cl

Sin embargo, pese a las diferencias observadas dentro del entorno de la ciudad, todavía existe un amplio debate sobre los diferentes procesos de caracterización de los grupos socioeconómicos que componen la ciudad. Esta situación tiene dos efectos relevantes: Por un lado, la amplio debate ha imposibilitado la posibilidad de poder definir la variable ha generado una proliferación de indicadores a partir de diferentes componentes, dificultando la capacidad de comparabilidad entre ellos. Por otra parte, la dificultad de poder establecer un indicador único impide establecer evaluaciones consistentes sobre el fenómeno de la desigualdad en la ciudad.

En este contexto, la necesidad de poder estratificar de manera sistemática a los diferentes grupos de la ciudad, de tal manera que pueda integrar los elementos económicos (dinero, vivienda) pero al mismo tiempo otros elementos relevantes asociados que también pueden estar influyendo en estos procesos de concentración (Educación, Tipo de trabajo, acceso a redes de poder etc) requiere de la elaboración de un modelo predictivo que permita establecer a partir de ciertas covariables un estimación aproximada de la variable de interés. Pero, para poder llevar a cabo dicha tarea se deberá solucionar los problemas asociados a la selección de un variable de interés que pueda ser medida y la necesidad de poder tener información sobre las covariables al menor nivel desagregado posible.

Frente a este escenario, la necesidad de poder caracterizar a los individuos que se están concentrando en diferentes espacios de la ciudad ha generado que la estratificación de estatus socioeconómico uno de los elementos relevantes a profundizar. Por un lado, algunas instituciones gubernamentales han establecido diferentes indicadores de caracterización social (MDS, INE, Método SAE), sin embargo, estas estimaciones no se han encontrado exentas de problemas. Estas soluciones también han sido expuestas a críticas, ya que las estimaciones se generan a través de atributos, de la respuesta generada por las propias familias, lo que dificulta la comparación y validación de los resultados obtenidos en distintos indicadores. Por otro lado, la utilización de variables externas también se ha presentado como otra alternativa aceptable, sin embargo, la dificultad de poder acceder a esta información hace aún más complicado la utilización de estas alternativas.

La gran cantidad de debates acerca del tema, se ha enfocado principalmente en análisis teórico del problema, simplificando la discusión del problema entorno a elementos de selección de variables de variables y su propiedades medición. En este sentido, el objetivo de esta investigación se orientará en poder establecer un análisis del indicador socioeconómico. A través de los diferentes elementos que conforman este indicador, se evaluarán las propiedades de medición del indicador y a través de sus diferentes usos, verificar la efectividad de estos elementos en la caracterización el fenómeno de estatus socioeconómico.

2 Objetivos

A partir de las conversaciones con la contraparte, el objetivo de este proyecto es generar un modelo de estatus socioeconómico. Asimismo, una vez generado dicho modelo, utilizando una homologación, se espera contrastar de manera gráfica la calidad predictiva del indicador.

Para poder llevar a cabo esta tarea, en este proyecto se desglosó en tres objetivos. Primero, recopilar y estandarizar diferentes fuentes de información con el fin de establecer un registro de información relevante que permita evaluar la selección de variables utilizadas que permita la posterior homologación. Segundo, evaluar la calidad explicativa del indicador socioeconómico. En tercer lugar, contrastar la efectividad de la escala dentro de un contexto geográfico; con el fin de poder evaluar los resultados del punto anterior, se contrastará la información obtenida en conjunto a diferentes variables relevantes de la encuesta nacional de caracterización social CASEN 2017, permitiendo comparar desde un punto de vista geográfico la capacidad explicativa del fenómeno del estatus social.

La realización de este objetivo también estarán asociados al desarrollo de ciertos objetivos específicos. La definición de un modelo predictivo sobre el ingreso permitirá la posibilidad de observar la existencia de patrones a nivel comunal. En este sentido, basado en información previa

2.1 Objetivo general

- Generar un modelo de estatus socioeconómico utilizando la encuesta CASEN 2017 utilizando como variable auxiliar (proxy) ingreso per cápita.

2.2 Objetivos específicos

- Identificar patrones geográficos de agrupaciones de valores predichos según información previa de comunas.
- Contrastar los resultados de los modelos predictivos de los modelos propuestos (multinomiales y de regresión múltiple).

3 Contextualización y descripción de la problemática

El estudio se centrará en la zona geográfica correspondiente a la Región Metropolitana, seleccionado por la contraparte, debido a la alta concentración de población que presenta y la alta desigualdad presente [4]. Utilizando los registros públicos de la encuesta CASEN, se generará un modelo predictivo con las variables seleccionadas por la contraparte con el fin de analizar la calidad del modelo junto con proporcionarnos información en las distintas fases del proyecto. En caso de que se requiera información con algún tipo de restricción, se le solicitará a

la contraparte que ayude con el proceso de validación de apoyo institucional requerido.

Asimismo, los programas e información requerida por la consultora son de carácter gratuito por lo que no se requerirá el apoyo de la contraparte en la utilización de licencias o pagos asociados para la utilización de estos programas.

3.1 Requerimientos

Los requerimientos de información, se detallan a continuación:

- Datos censales 2017: base de datos de personas, hogares y viviendas (información relacional), así como de las manzanas y macromanizas censales (información espacial).
- Datos CASEN 2017: Esta base de datos contiene información acerca de todos los elementos previamente seleccionados.

Una vez entregado cada uno de los insumos previamente establecidos, se le pedirá a la contraparte que pueda establecer un informe de resultados en el cual pueda señalar de manera explícita las modificaciones necesarias.

3.2 Análisis Preliminar

Durante la última década la ciudad de Santiago ha sufrido un aumento explosivo en sus habitantes. El censo del año 2017, registró un aumento en un 18% en comparación del censo anterior (2002), llegando a registrar más de 7 millones de habitantes. Este crecimiento sistemático en la ciudad ha estado marcado dentro del contexto latinoamericano el cual ha presenciado un importante crecimiento de las áreas urbanas y particularmente las metrópolis, generando drásticos cambios en su conformación. Kloosterman y Musterd (citado en Truffello & Hidalgo, 2015)¹ plantean que este proceso a estado asociado a la diversos factores, entre los cuales cabe destacar la explosión demográfica, los procesos de suburbanización, los flujos migratorios y la desconcentración de los procesos productivos de los países han ayudado a establecer este pararama en las urbes latinoamericanas.

Este aumento en la población no ha sido inocuo en la ciudad, la gran cantidad de nuevos habitantes en la metrópolis ha generado un fuerte impacto en los valores de los terrenos, la desregularización de las políticas y el impulso por parte del gobierno en la urbanización de los sectores periféricos del gobierno han traído importantes cambios en la ciudad de Santiago.

El creciente incremento en la cantidad de habitantes en la ciudad ha generado una explosión en la demanda por viviendas, lo que ha motivado el interés por la compra y subarriendo de propiedades, generando la disminución sostenida de viviendas orientadas a personas de bajos ingresos e influyendo

¹ 2

directamente en la diferenciación de sectores por su capacidad de pago y en las posibilidades de movilidad dentro de la ciudad.²

La liberación de los mercados urbanos en Santiago basado en los cambios económicos impulsados en los años ochenta en Chile tuvieron como principal consecuencia la concentración de capital a través de la unión de todos los aspectos del proyectos urbanos, permitiendo que los procesos de adquisición de terreno, diseño, urbanización, financiamiento y publicidad en un mismo actor, permitiendo definir los proyectos urbanos entorno a intereses particulares.

Durante los años ochenta, se planteo como prioridad la disminución al déficit de viviendas, lo que definió gran parte de las políticas públicas sobre vivienda hasta finales del los años noventa. Uno de los principales manifestaciones de esta tendencia fue la compra por parte del estado de grandes paños de terrenos en zonas periféricas de la ciudad para la construcción de bloques de viviendas. Estos proyectos orientados solo en la necesidad de erradicar los déficit de viviendas, tuvieron nosivos impactos en la conformación de la ciudad en posteriores décadas. La erradicación de campamentos y conformación de sectores de la ciudad alejados de servicios y conexiones interurbanos, han generado negativos efectos en la calidad de vida de un sector de la población.

Santiago ha sido parte de grandes proceso sociales durante las últimas décadas, los importantes cambios demográficos, económicos y urbanos de los que ha generado también la necesidad de poder analizar con mayor profundidad los diferentes fenómenos que están ocurriendo en la ciudad. Sin embargo, este interés no se encuentra exento de problemas metodológicos. La gran cantidad de disciplinas interesadas han diversificado la definición de los fenómenos urbanos, generando confusión conceptual, discutiendo sobre un mismo fenómeno argumentado por diferentes corrientes, impidiendo la posibilidad de comparar efectos y resultados.

En este contexto, la segregación urbana ha sido un elemento de amplio debate académico. Los altos índices de concentración de riqueza que se han observado en el último tiempo, en conjunto con los cambios que ha vivido la ciudad durante las últimas décadas han vuelto cada vez más relevante la necesidad de poder conocer cómo se están estructurado la urbe.

El fenómeno de la segregación, el cual busca caracterizar las diferencias sociales dentro de un plano geográfico acotado, se ha visto enfrentado dos problemas relevantes. Por un lado, la dificultad de poder definir un método para la estimación de grupos sociales que no se encuentre exento de debate. Por un lado, la necesidad de poder estratificar requiere de definir un mecanismo que pueda incluir los diversos factores económicos, sociales y culturales que caracterizan el fenómeno de estatus social. Frente a esta situación, la aproximación a este fenómeno se ha centrado en los estudios de los niveles de segregación de ingreso, donde la variable ingreso se presenta como una opción de mayor acceso y además permite la posibilidad de poder comparar con otros países y verificar los efectos de políticas públicas relacionadas (OCDE,2018).

² 3

Por otro lado, tal como retrata Agnostini y Brown (2007) la posibilidad de graficar los mapas de desigualdad de ingresos pueden entregar una descripción de los fenómenos que se están investigando. Sin embargo, la falta de datos suficientemente detallados y desagregados imposibilitan en parte esta situación. Frente a este fenómeno, se ha generado la necesidad de estimar modelos que puedan alinear la información a gran escala a nivel desagregado a través de la estimación de modelos de imputación de datos, a partir de otras variables estimar información acerca del ingreso promedio.

4 Metodología

A través de las aproximaciones del fenómeno de la segregación a través de los estudios de ingreso, se ha propuesto la construcción de modelos que permitan generar valores aproximados de ingreso de cada persona. Usando datos de encuestas de hogares para generar valores que serán calculados en base a la información del censo, estimando indicadores de ingreso a nivel desagregado. Para ello se utilizó en este caso como variable de a predecir la variable ingreso per cápita por hogar la cuál corresponde a la cantidad de ingreso por hogar dividido en el número de personas que hay por hogar. Posteriormente se utilizan usando los datos correspondientes a variables de observables que se encuentren en ambas encuestas (CASEN-CENSO 2017) con el fin de establecer alguna posterior homologación. Una vez estimado el modelo, usando los parámetros estimados en la primera parte, se genera un valor predicho para cada uno de los casos del CENSO, los cuáles posteriormente serán graficados.

Estadísticamente, la metodología consiste en estimar una distribución conjunta de la variable de interés en conjunto con un grupo de covariables. En la primera parte se generó un modelo de regresión múltiple con una serie de variables:

$$\ln(y_{pcln}) = \beta_0 + \beta_1 educ + \beta_2 sexo + \beta_3 edad + \beta_4 n.piezas + \beta_5 tot.hog + \epsilon$$

con un $\epsilon(0, \Sigma)$.

En la segunda parte, una vez generado el modelo multinomial se plantea generar un modelo alternativa, esta vez constituido por un modelo multinomial para predecir el ingreso:

$$\begin{aligned} \text{Categorías}(y_{pcln}) = & \beta_{11} + \beta_{11}(y_{pcln} = 2) + \beta_{12}(y_{pcln} = 3) + \beta_{13}(y_{pcln} = 4) + \\ & \beta_{14}(y_{pcln} = 5) + \beta_{15} sexo + \beta_{16} edad + \beta_{17} n.piezas + \\ & \beta_{18} tot.hog + \epsilon \end{aligned}$$

la encuesta Casen es realizada por el Ministerio de Desarrollo Social (MDS), pero los datos son posteriormente ajustados por la Comisión Económica para América latina y el Caribe (Cepal) usando el sistema de cuentas nacionales como referencia. Estos ajustes consideran los problemas generados por la falta

de datos de ingreso para algunos hogares y la sub o sobrerrepresentación de algunas categorías de ingreso en la muestra. la encuesta utiliza un método de muestreo aleatorio de etapas múltiples con estratificación. En la primera etapa, el país se divide entre las áreas rurales y urbanas para cada una de las 20 regiones. las unidades de muestreo primario son seleccionadas con probabilidades proporcionales a la población. En la segunda etapa, una vez que la estratificación está hecha, se seleccionan hogares de tal forma que todos tengan la misma probabilidad de ser elegidos en la muestra.

El instituto nacional de Estadísticas realiza un censo de población y vivienda en Chile cada 10 años. El censo utilizado en este trabajo fue realizado el 2017, por lo que no hay distancia temporal respecto a la encuesta CASEN.

4.1 Variables disponibles entre CASEN/CENSO

Los datos ya vienen depurados e imputados, por lo que no es necesario hacer mayores revisiones a los mismos.

Las observaciones de datos disponibles inicialmente para la RM corresponden a 42.601 (804 variables) y 7.022.192 (53 variables), para CASEN y CENSO respectivamente.

Las siguientes variables son las que hasta el momento se han encontrado en común entre CASEN y CENSO. Las variables *ypc* e *ypcln* son las únicas que no están en el CENSO, que es en el fondo la variable respuesta que se pretende calcular (primero armar un modelo usando CASEN y luego ocupar el modelo con los datos censales).

Variable	Descripción	Tipo
<i>ypc</i>	Ingreso Total per cápita del hogar corregido (CASEN 2017)	Continua
<i>ypcln</i>	$\log[\text{ypc} + 1]$	Continua
<i>educ</i>	Nivel educacional (según la definición de CASEN 2017)	Catagórica
<i>esc</i>	Años de escolaridad (CASEN y censo)	Continua
<i>sexo</i>	Hombre/Mujer (CASEN y censo)	Catagórica
<i>edad</i>	Años, de 0 a 100 (según la definición del censo 2017)	Continua
<i>n_piezas</i>	Número de habitaciones por vivienda (según censo 2017, que tiene un máximo de 6 habitaciones)	Continua
<i>comuna</i>	Código de comuna (CASEN y censo)	Catagórica
<i>tot_per</i>	Total de personas por hogar (CASEN y censo)	Continua
<i>tot_hog</i>	Total de hogares por vivienda (CASEN y censo)	Continua
<i>en_mapa</i>	Comunas que se despliegan en el mapa	Binaria

Los datos pertenecen sólo a la RM (completa). La variable *en_mapa* es simplemente un flag que indica si dicho registro será desplegado en el mapa (en el mapa, no se muestra toda la RM por razones prácticas, se muestran las de la Provincia de Santiago y Puente Alto).

4.2 Homologación

Para todas las variables censales, los registros que contaban con 98 (No Aplica) y 99 (Missing) como valor observado, se transformaron en NA.

esc (ESCOLARIDAD), sexo (P08), comuna (COMUNA), tot_hog (CANT_HOG) y tot_per (CANT_PER) son variables que no requieren transformación entre ambos conjuntos de datos y se pueden utilizar indistintamente puesto que cada codificación representa lo mismo (Los valores entre () denotan el nombre de la variable en la base de datos original del CENSO).

La variable edad (P09) fue modificada en CASEN, para que calzara con el máximo registrado por el CENSO; de esta forma las edades mayores a 100, se agruparon y pasaron a formar parte de la categoría 100 (al igual que el CENSO).

La variable n_piezas (P04) es un caso similar; el número de piezas máximas registradas por CENSO es de 6, mientras que en CASEN no hay un límite. Por dicha razón, se hizo una recodificación de CASEN, agrupando en el valor 6 o más, al igual que el CENSO.

5 Resultados del análisis exploratorio

La tabla 1 muestra las métricas habituales para las variables continuas que fueron seleccionadas desde CASEN 2017. La variable de interés a explicar, ingreso per cápita (ypc), posee una amplitud muy grande en cuanto a los valores que puede tomar, teniendo sus valores promedio y su mediana, muy alejados del valor máximo (lo cual está en línea con los reportes sobre la desigualdad del ingreso que normalmente se divulgan).

Table 1 Métricas básicas de variables explicativas continuas

	ypcnc	ypcc	esc	edad	n_piezas	tot_per	tot_hog
Mínimo	0	0	0	0	0	1	1
1° Cuantil	4,6	42.629	10	19	2	3	1
Mediana	4,9	83.450	12	34	3	4	1
Media	5,0	192.427,3	11,9	36,8	3,1	4,0	1,2
3° Cuantil	5,3	187.083,5	15	54	4	5	1
Max	7,5	30.566.664	22	100	6	19	10
NA's	0	0	8.365	0	0	1	1

Algunos datos interesantes (también en línea con otros reportes) es que en promedio en la RM, los hogares están constituidos por 4 personas y la escolaridad alcanza los 12 años (equivalente a 4° medio).

En la figura 1 se pueden apreciar como se comportan las variables explicativas educación, escolaridad, sexo y edad, con respecto a las variables de interés ingreso y $\log(\text{ingreso})$. De inmediato se puede notar que al revisar la relación entre ingreso y $\log(\text{ingreso})$, este último permite revisar

de mejor manera cualquier tipo de interacción con la variable explicativa, lo cual se produce por la alta concentración que tiene esta variable (muy pocas personas ganan mucho, y muchas personas ganan poco).

Para estas 4 variables explicativas, no se logra apreciar ningún tipo de relación de interés para $\log(\text{ingreso})$, salvo quizás por la educación y escolaridad, que muestran una muy leve tendencia a incrementar, conforme aumenta el nivel educacional.

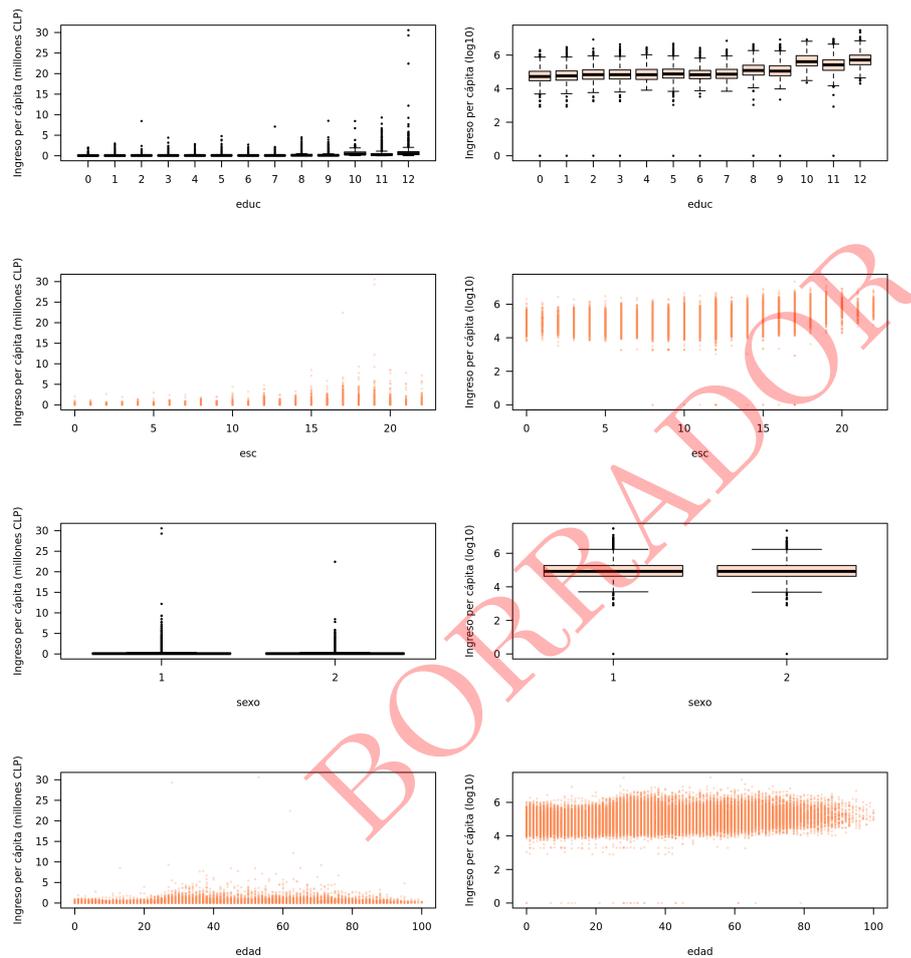


Fig. 1 Variables respuestas versus predictororas: educación, escolaridad, sexo y edad.

En la figura 2 se pueden apreciar como se comportan las variables explicativas número de piezas, comuna, personas por hogar y hogares por hogar,

con respecto a las variables de interés `ingreso` y `log(ingreso)`. El comportamiento de estas 4 variables es similar en cuanto a que `log(ingreso)` presenta una mayor facilidad para apreciar la relación con las respectivas variables respuesta.

BORRADOR

Comparando contra `log(ingreso)`, se puede apreciar que la variable categórica de comuna, tiene algunas categorías que presentan saltos y bajos importantes con respecto a la tendencia general. A su vez, el número de piezas y el total de hogares por vivienda tampoco presentan una relación lineal con la variable de interés. No obstante, se puede apreciar que el total de personas por hogar sí tiene una tendencia lineal negativa bastante más marcada que las otras covariables; es decir, conforme aumenta el número de personas en el hogar, el ingreso per capita tiende a disminuir (lo cual tiene bastante sentido).

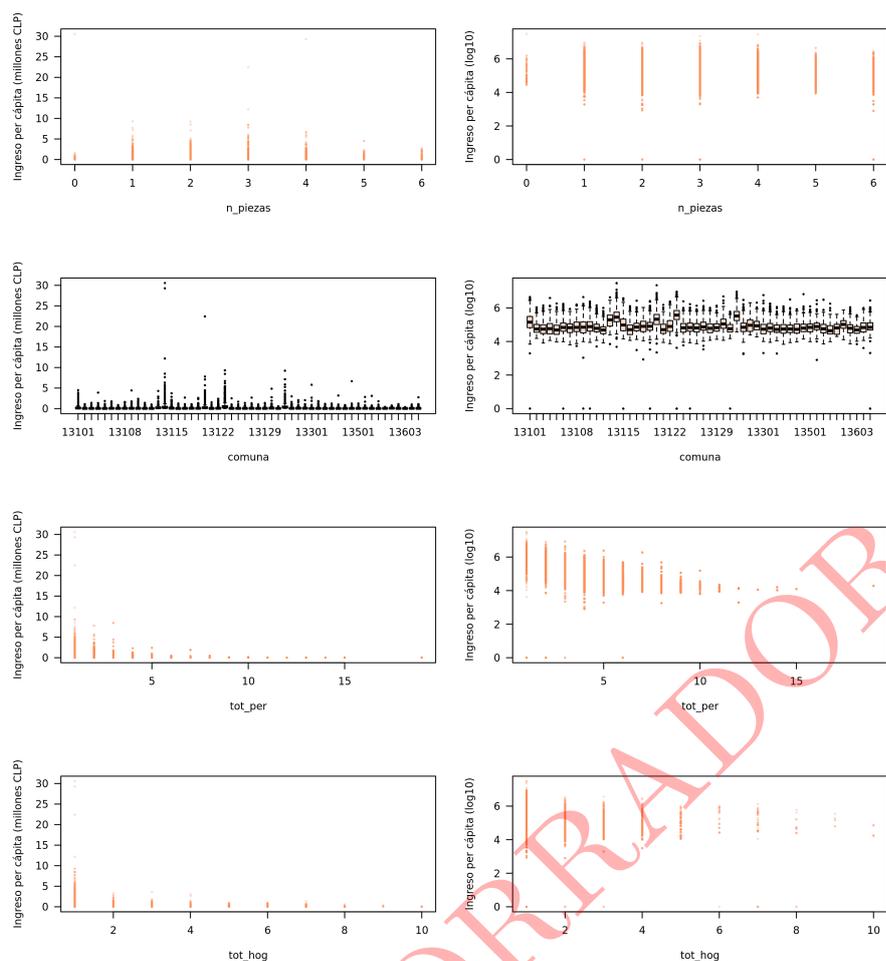


Fig. 2 Variables respuestas versus predictoras: número de piezas, comuna, personas por hogar y hogares por vivienda.

6 Modelo y Validación del modelo

6.1 LM

El modelo lineal se ajustó con las variables educ, sexo, edad, n_piezas y tot_hog.

En el modelo ajustado (tabla 2). Se observa que la mayoría de la variables tienen un aporte positivo en la predicción del ingreso, a excepción del sexo y el número de piezas.

Por otra parte, la gran mayoría de la variables es significativa; aquellas que no lo son, corresponden a dos categorías de la variable educ. Sin embargo, no

son una gran cantidad de categorías, por lo cual, no corresponde eliminar la variable en su totalidad.

Además, se obtiene un R^2 ajustado del 31,4%, un AIC de 87.719,3 y un BIC de 87.871,05.

Respecto a los supuestos del modelo (figura 3), se aprecia que existe un linealidad de la media adecuada; respecto a la heterocedasticidad de los errores, teniendo una gran cantidad de observaciones, se aprecia una amplitud constante a medida que aumenta el valor ajustado. Finalmente, respecto al supuesto de normalidad de los errores, se observa una cola superior bastante pesada, la cual puede ser indicio de que se rechaza la hipótesis nula.

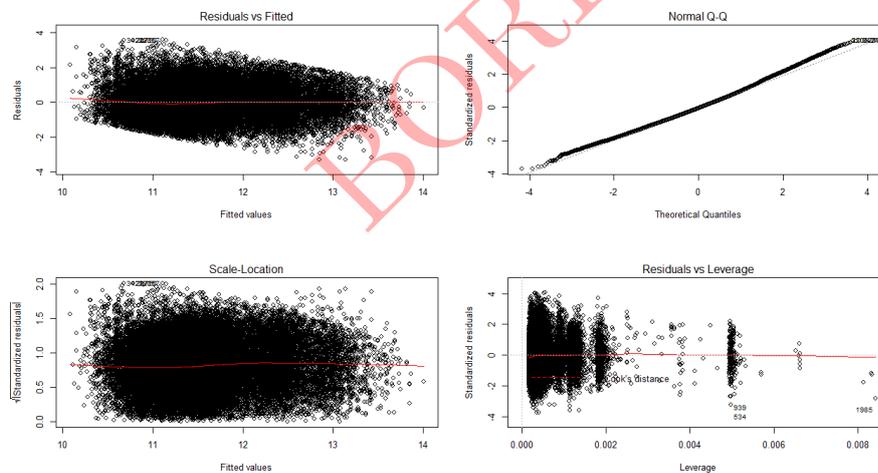


Fig. 3 Distribución de los errores del modelo lineal

Table 2 Cuadro resumen del modelo lineal ajustado.

	<i>Dependent variable:</i>
	ypcln
educ1	0.023 (0.041)
educ2	0.064 (0.041)
educ3	0.294*** (0.040)
educ4	0.293*** (0.048)
educ5	0.374*** (0.039)
educ6	0.330*** (0.041)
educ7	0.587*** (0.046)
educ8	0.855*** (0.041)
educ9	1.021*** (0.041)
educ10	2.050*** (0.072)
educ11	1.514*** (0.039)
educ12	2.109*** (0.047)
sexo2	-0.047*** (0.010)
edad	0.017*** (0.0003)
n_piezas	-0.134*** (0.004)
tot_hog	0.115*** (0.009)
Constant	10.502*** (0.044)
Observations	33,877
R ²	0.314
Adjusted R ²	0.314
Residual Std. Error	0.883 (df = 33860)
F Statistic	969.855*** (df = 16; 33860)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

6.2 GLM

Un segundo modelo planteado es un modelo multinomial, en el cual se busca agrupar los ingresos por percentiles.

El modelo se ajustó con las mismas variables utilizadas en el modelo lineal, además, la variable respuesta ($\log(\text{ingreso})$), está categorizada bajo los percentiles (20, 40, 60 y 80), generando así, cinco posibles rangos (tabla 3). Los resultados del ajuste se pueden apreciar en la tabla 5. En general, se puede apreciar que la educación tiene un efecto positivo, al igual que la edad. Por otro lado, el número de piezas por vivienda y ser mujer, tiene un efecto negativo en el ingreso per cápita.

Table 3 Categorías a predecir por el modelo multinomial.

Categoría	1	2	3	4	5
Percentil	0-20%	20-40%	40-60%	60-80%	80-100%
CLP (miles)	< 36,82	63,89	109,6	226,46	> 226,46

Por otra parte, en la tabla 4 se pueden observar los valores resultantes del análisis para la matriz de confusión del modelo (con un Accuracy = 33,65%).

Table 4 Tabla resumen

Nivel	1	2	3	4	5
Precision	57.59%	15.49%	12.07%	30.10%	52.92%
Recall	34.35%	25.61%	24.56%	28.04%	45.90%
Specificity	83.20%	63.18%	60.92%	71.71%	77.52%

Finalmente el AIC del modelo arroja un valor de 122.743,3 y un BIC igual a 123.331,1.

Table 5 Tabla resumen modelo GLM ajustado

	<i>Dependent variable:</i>			
	2 (1)	3 (2)	4 (3)	5 (4)
educ1	-0.006 (0.059)	-0.097 (0.064)	-0.297*** (0.070)	-0.741*** (0.075)
educ2	-0.098 (0.078)	-0.216*** (0.081)	-0.502*** (0.085)	-1.130*** (0.092)
educ3	0.070 (0.069)	0.034 (0.072)	-0.229*** (0.077)	-0.722*** (0.083)
educ4	-0.133 (0.119)	-0.204* (0.123)	-0.342*** (0.126)	-0.800*** (0.138)
educ5	0.190*** (0.065)	0.151** (0.068)	-0.064 (0.072)	-0.510*** (0.076)
educ6	0.082 (0.078)	0.083 (0.080)	-0.262*** (0.086)	-0.906*** (0.097)
educ7	0.207** (0.102)	0.328*** (0.103)	0.031 (0.112)	-0.194 (0.119)
educ8	0.412*** (0.099)	0.747*** (0.097)	0.811*** (0.098)	0.791*** (0.099)
educ9	0.513*** (0.085)	0.856*** (0.085)	1.026*** (0.087)	1.117*** (0.088)
educ10	0.078 (0.709)	0.932 (0.603)	1.987*** (0.539)	3.248*** (0.515)
educ11	0.752*** (0.115)	1.384*** (0.110)	1.938*** (0.108)	2.564*** (0.106)
educ12	0.624 (0.559)	1.777*** (0.487)	2.892*** (0.462)	4.381*** (0.454)
sexo2	-0.064** (0.033)	-0.144*** (0.033)	-0.127*** (0.034)	-0.123*** (0.036)
edad	0.010*** (0.001)	0.018*** (0.001)	0.031*** (0.001)	0.037*** (0.001)
n_piezas	-0.101*** (0.015)	-0.124*** (0.015)	-0.195*** (0.015)	-0.239*** (0.016)
tot_hog	0.055* (0.028)	0.047 (0.029)	0.028 (0.032)	0.105*** (0.033)
Constant	-0.120* (0.068)	-0.323*** (0.072)	-0.536*** (0.077)	-0.691*** (0.081)
Akaike Inf. Crit.	114,468.000	114,468.000	114,468.000	114,468.000

Note:

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

7 Resultados

Los resultados de los modelos utilizados fueron presentados anteriormente en las descripciones y análisis de los modelos específicos. Si bien estos poseen características diferentes la comparación de sus AIC son de 91.260,68 para el modelo de regresión lineal y 122.743,3 para el modelo multinomial.

Junto a lo anterior es posible observar que debido a la gran cantidad de variables categóricas queda pendiente la posibilidad de poder probar algunas uniones de categorías para observar si existen diferencias relevantes.

7.1 Mapas de ingreso

En el mapa de ingreso que se presenta en la figura 4, se pueden apreciar los valores predichos a través del modelo de regresión lineal para la región metropolitana, donde es posible observar una concentración de valores altos de ingresos dentro del sector oriente de la ciudad, con la excepción de conos locales relevantes como es el caso del Cerro 18 en el sector de Lo Barnechea.

Esta situación muestra que si bien es posible caracterizar los sectores de menores recursos de la población pero las diferencias más grandes requerirán la incorporación de otras variables relevantes.

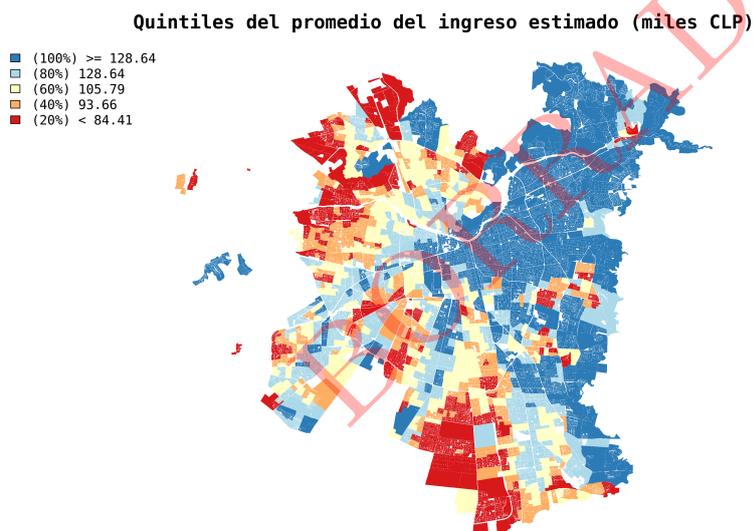


Fig. 4 Quintiles de la distribución promedio del ingreso en la RM

Distinta es la situación observada en la figura 5 en la cual se aprecia una distribución mucho más heterogénea del mapa de distribución de ingreso a la hora

de utilizar una nueva tipificación de 8 cuantiles con especial interés en diferenciar los sectores más altos de la población - Específicamente, tres cuantiles están diseñados para la separación de los segmentos 90,95 y 99 de la población -. En este sentido, es posible observar ciertas diferencias más acentuadas en ciertos sectores de la ciudad, donde pese a que se mantiene el porcentaje de riqueza dentro de los sectores de Vitacura, Providencia y Las Condes, este se ve disuelto en el caso de las comunas de La Reina y Lo Barnechea.

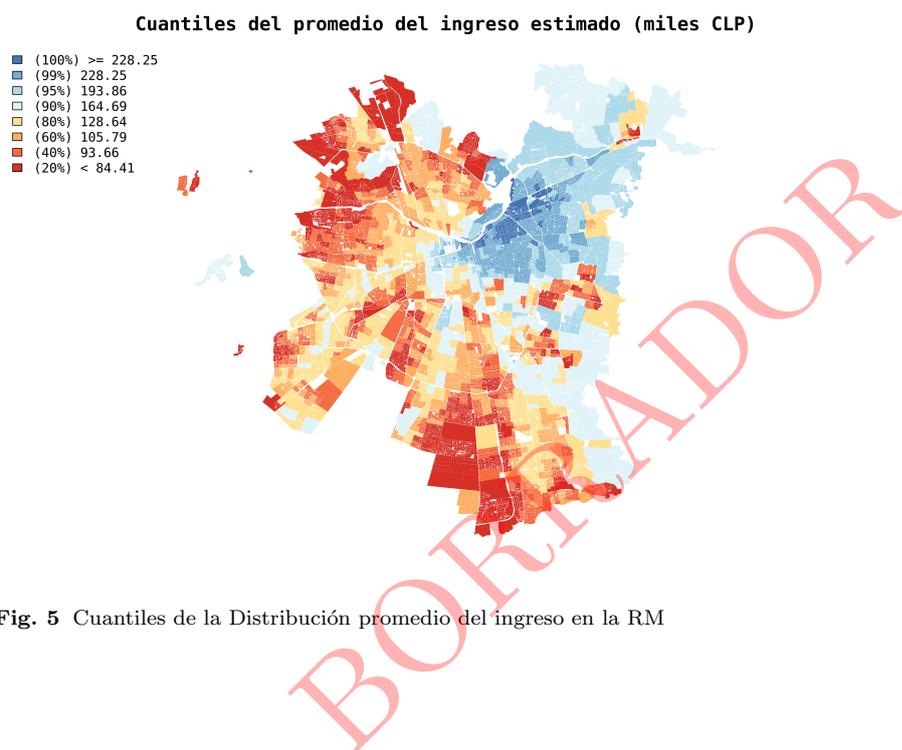


Fig. 5 Cuantiles de la Distribución promedio del ingreso en la RM

Similar a lo anterior es lo observado cuando modelamos los quintiles de la mediana del ingreso estimado (figura 6). La utilización de esta partición se ve enfrentada a los problemas similares que habían sido observados anteriormente en el proceso de graficar los resultados de la media. La alta concentración de recursos hace poco plausible poder discernir entre los grupos más acomodados de la población, los que no sugiere la necesidad de generar modelos para grupos específicos o la inclusión de nuevas variables al modelo.

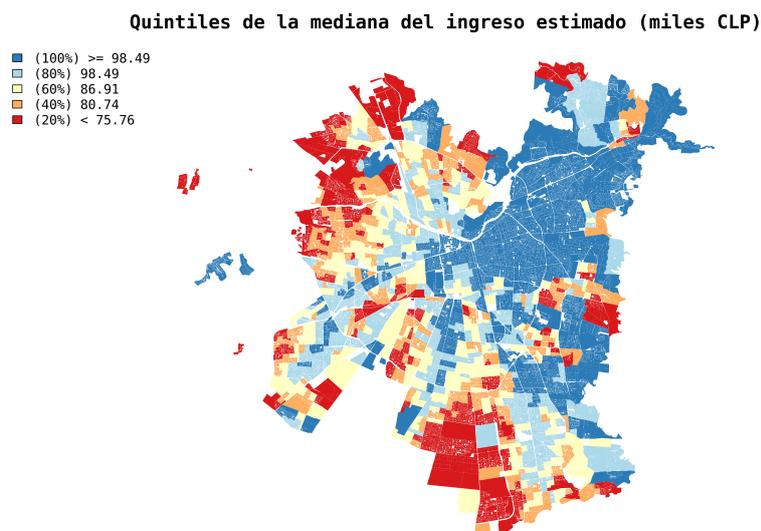


Fig. 6 Quintiles de la Distribución de la mediana del ingreso en la RM

No obstante, la figura 7 correspondiente a los cuantiles para la mediana predicha, es posible observar que existen importantes diferencias con el gráfico observado en la media, donde es posible observar importantes cambios en el color de algunos sectores. Gran parte de estos terrenos que poseen grandes extensiones de terreno también son indicadores de importantes diferencias de ingresos dentro de estos terrenos lo que también puede ser un interesante sector a analizar en profundidad ya que este puede entregar valiosa información con respecto contacto entre habitantes de diferentes ingresos.

Por último se puede apreciar la figura 8 correspondiente a la información obtenida a través del modelo multinomial. Cómo es posible observar se puede apreciar un patrón mucho menos organizado a los observados anteriormente. En este sentido, la distribución del ingreso presenta una mayor varianza debido a que es estimada por cada habitante lo que permite observar que dentro de un mismo grupo familiar pueden existir diferentes rangos de ingreso. En este sentido, la gran cantidad de información que se genera por manzana termina teniendo repercusiones negativas en la interpretación del modelo.



Fig. 7 Cuantiles de la Distribución de la mediana del ingreso en la RM

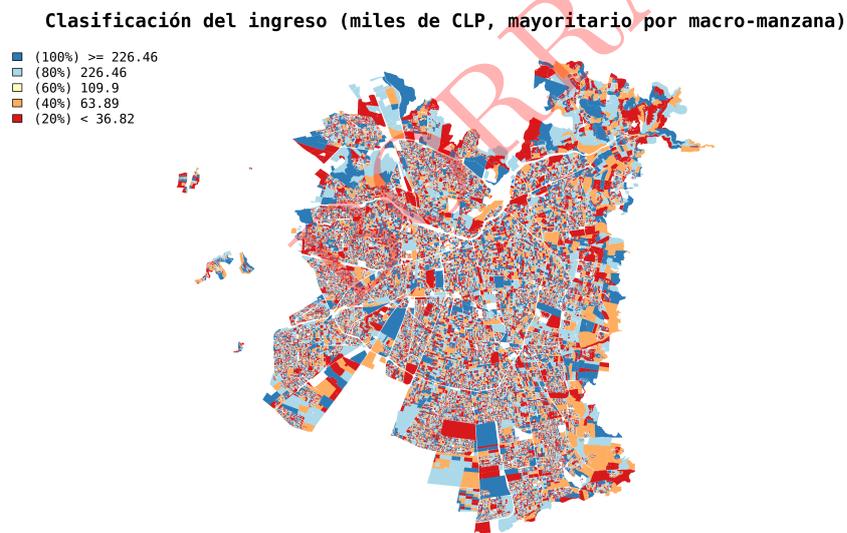


Fig. 8 Categorías de la distribución ingreso en la RM

8 Conclusiones

La segregación dentro de Latinoamérica se ha vuelto un tema cada vez más relevante dentro de la discusión a nivel de políticas públicas y social. Chile no se ha encontrado fuera de esta situación, donde los sistemáticos crecimientos que observó el país durante los años noventa se tradujeron en importantes cambios en la conformación de la ciudad.

Sin embargo, las importantes diferencias de ingreso y desigualdad que se han empezado a observar en la ciudad (Sabatini, y Salcedo, R. 2007) han presentado cada vez más dificultades metodológicas para poder definir y estimar fenómenos. Específicamente, el fenómeno de la segregación se ha enfrentado a estos problemas.

En este sentido, los modelos presentados han permitido bajo el alero de los modelos de estimación de ingreso y imputación de datos proporcionar un modelo que presenta resultados esperados sobre la distribución de ingreso del país. Sin embargo, la poca capacidad explicativa de los modelos propuestos (R^2 cercano al 31%) y los problemas observados en los residuos hacen difícil poder quedar satisfechos con este tipo de modelos.

En primer lugar, vuelve a emerger el problema asociado a la información que se posee para la estimación del modelo. Debido a que el fenómeno de las diferencias sociales solo ha quedado simplificado a el nivel ingreso de los habitantes, este ignora elementos relevantes, como ingreso intergeneracional y otros aspectos sociales como calidad de vida, redes sociales y recursos culturales, los cuáles tampoco pueden ser solamente definidos entorno a la capacidad de recursos. En este sentido, se vuelve imperioso complejizar el fenómeno de interés de tal manera que pueda contemplar la complejidad del fenómeno a estudiar.

Segundo, la posibilidad de que estos eventos se encuentren asociados de manera teórica en un posible concepto de estatus socioeconómico, genera un conflicto a la hora de modelar el ingreso utilizando estas covariables. Específicamente, si se considera que el nivel educacional es un aspecto relevante para caracterizar el estatus alcanzado por una persona, se hace necesario repensar el rol del nivel educativo a ser parte de la variable de interés.

Tercero, uno de los modelos contemplados fue la incorporación de la variable comuna al modelo. Pese a que este generó amplias mejoras al modelo este afectaba la parsimonia del modelo. Sin embargo, cabe destacar la fuerte relevancia que posee la comuna a la hora de predicción en el ingreso, lo que puede sugerir la necesidad de profundizar en mayor detalle sobre la implementación de modelos mixtos (colocando a la comuna como conglomerado) y la búsqueda de información a nivel comunal la cuál pueda entregar mayor valor predictivo al modelo.

Cuarto, la generación de modelos de ingreso para los habitantes a través de la imputación no nos ha dejado indiferente frente al valor que poseen los errores en las estimaciones puntuales. Considerando la relevancia de efectos comunales, existe la posibilidad de que las estimaciones puntuales generadas posean errores no identificados, por lo que se plantea la necesidad de analizar estos resultados con cautela y a modo descriptivo, ya que mientras no sea

posible descartar dichos efectos, la subestimación de la varianza se plantea como un escenario altamente probable.

En definitiva, la segregación dentro de la ciudad de Santiago es un fenómeno de alta complejidad de análisis. Frente a este escenario complejo, la estimación de modelos de ingreso se ha presentado como una opción de fácil acceso y la cuál en consistencia con información externa ha podido caracterizar de manera sistemática las diferentes concentraciones de recursos dentro de la ciudad. Sin embargo, los resultados obtenidos en este trabajo ponen en duda la veracidad de las conclusiones que se bajaran utilizando estos modelos, donde la utilización de estos modelos implicará además la necesidad de poer incorporar la utilización de nuevas variables o la inclusión de nuevos modelos de medición que permitan descartar problemas asociadas a la inferencia de los resultad obtenidos con estos.

9 Próximos Pasos

- Volver a profundizar en la posibilidad de integrar elementos que geográficos o información externa para la elaboración de indicadores. Considerando la dificultad para abarcar temas como recursos educativos y redes sociales, es necesario volver a mirar información bibliográfica con el fin de poder determinar posibles variables que puedan ayudar a caracterizar estos fenómenos a cabalidad.
- Analizar la posibilidad de implementar modelos mixtos, los cuáles puedan incluir posibles efectos temporales (i.e recursos heredados a través del tiempo) o de estratos (i.e efectos asociados a sectores de la ciudad o comunas).
- Plantear otras formas de modelar el nivel socioeconómicos, caracterizando la variable de otras formas (i.e Acceso a recursos específicos o la posibilidad de acceso a elementos en particular).
- Profundizar en los estratos generados con el fin de poder generar perfiles de las macro-manzanas que serán modeladas

10 Bibliografía

- [1]https://www.nytimes.com/2014/12/03/business/economy/income-gap-shrinks-in-chile-for-better-or-worse.html?_r=0
- [2]<https://www.economist.com/the-americas/2012/04/14/progress-and-its-discontents>
- [3]<https://repositorio.uc.cl/bitstream/handle/11534/3904/000335142.pdf?sequence=1>
- [4]Sabatini, F. (2006). La segregación social del espacio en las ciudades de América Latina. Inter-American Development Bank.
- [5]Sabatini, F., & Salcedo, R. (2007). Gated communities and the poor in San- tiago, Chile: Functional and symbolic integration in a context of aggres-

sive capitalist colonization of lower-class areas. *Housing Policy Debate*, 18(3), 577-606. <http://doi.org/10.1080/10511482.2007.9521612>

[6] Mueller, C., & Parcel, T. (1981), Measures of Socioeconomic Status: Alternatives and Recommendations. *Child Development*, 52(1), 13-30.

[7] Blau, P. M., & Duncan, O. D. *The American occupational structure*. New York: Wiley, 1967

[8] OECD (2018), *Divided Cities: Understanding Intra-urban Inequalities*, OECD Publishing, Paris, <https://doi.org/10.1787/9789264300385-en>.

[9] Agostini, C., & Brown, P. (2007). *Desigualdad geográfica en Chile*.

BORRADOR