

I. INTRODUCCIÓN

Durante el periodo de rápido crecimiento que el país ha experimentado en la última década, los niveles de pobreza han disminuido considerablemente, tanto en el área rural como en el área urbana. Así, entre 2004 y 2015 la incidencia de pobreza en el área rural disminuyó en 38,3 puntos porcentuales mientras que la pobreza urbana cayó en 33,7 puntos. En 2015, la tasa de pobreza rural se sitúa en 45,18%, la urbana de 14,54% y la de Lima metropolitana es de 10,95%. Al mismo tiempo, la población de las principales ciudades ha continuado expandiéndose a un ritmo superior al de la población rural; tanto por el crecimiento demográfico como por las migraciones internas. El porcentaje de población rural ha caído de 29,9% en 2004 a 23,3% en 2015 de suerte que la población urbana representa hoy en día más de tres cuartos de la población nacional (76,7%). Ello ha implicado por un lado que el número de hogares en situación de pobreza sea en 2015 ligeramente superior al número de pobres en el área rural (3'517 446 pobres urbanos y 3'375,677 pobres rurales) y por el otro lado que sea cada vez más difícil estimar con la misma precisión la incidencia de pobreza cuando éstas alcanzan bajos niveles. Este problema es particularmente agudo en la capital en donde reside casi un tercio de la población (31,6% en 2015). Esta evolución de la pobreza ha significado la necesidad de tomar en cuenta la heterogeneidad de condiciones de vida de la población urbana en donde los hogares en situación de pobreza tienden a concentrarse en ciertas áreas de la ciudad.

Esta constatación ha motivado la necesidad de disponer de nuevos instrumentos que permitan identificar los lugares de la ciudad en donde se concentra la población en situación de pobreza. Ello posibilitará también establecer las distintas dimensiones de privaciones de dicha población. Los resultados obtenidos permitirán responder a las exigencias de información de las nuevas estrategias de lucha contra la pobreza y de la política social poniendo énfasis en las especificidades de la pobreza urbana. El poder distinguir diferentes áreas de la ciudad en donde se concentra la pobreza en sus diferentes dimensiones permitirá una intervención focalizada y con información actualizada.

Un ejemplo reciente de la creciente demanda de información estadística en áreas menores, es la política de subsidios según la clasificación socioeconómica que se sustenta en el artículo 10º del Decreto Supremo N°021-2012-EM, Reglamento de la Ley N°29852 y la Resolución Ministerial N°262-2016-MEM/DM, Programa de Masificación del Uso Residencial y Vehicular del Gas Natural.

En este documento se presenta la estratificación de las manzanas en el área urbana de Lima Metropolitana en función de los ingresos estimados de los hogares, permitiendo distinguir al interior de la ciudad, zonas con mayor o menor nivel de ingreso. Para ello se han clasificado los niveles de ingreso en 5 estratos: Alto, Medio Alto, Medio, Medio Bajo y Bajo, mediante procedimientos estadísticos que optimizan la homogeneidad intra estrato y maximizan las diferencias entre estratos, lo cual implica un menor error de focalización.

Para la estimación de ingresos por manzanas se ha empleado la metodología de estimación en áreas menores pues las encuestas por muestra no permiten estimaciones a ese nivel de desagregación. Los censos de población no pueden captar información sobre los ingresos de los hogares por ser un indicador complejo, más aún cuando los ingresos informales predominan. Por otro lado, tenemos las encuestas especializadas que si tienen obtienen información de ingresos de forma robusta a través de módulos especializados. Sin embargo, el diseño de dichas encuestas (como la Encuesta Nacional de Hogares) permite estimar indicadores de ingresos de los hogares con un nivel de representatividad solo departamental o para el conjunto de Lima metropolitana.

La metodología de estimación en áreas menores consiste en combinar la información de una encuesta (Encuesta Nacional de Hogares 2012-2013) con la de un censo (Empadronamiento Distrital de Población y Vivienda 2012-2013), además de otras fuentes de datos para estimar el ingreso per cápita para unidades administrativas menores (manzanas). En poder utilizar múltiples fuentes de datos nos permite captar con mayor precisión la heterogeneidad espacial de los niveles de ingreso. La metodología empleada sigue los mismos lineamientos que fueron implementados en la estimación del mapa de pobreza distrital 2013 publicado este año por el INEI. El documento que presentamos permite visualizar a nivel de cada distrito de Lima Metropolitana, las manzanas agrupadas en estratos según sus niveles estimados del ingreso per cápita del hogar.

II. ANTECEDENTES

2.1. Anterior estratificación de manzanas según el ingreso estimado

En el año 2009, el Instituto Nacional de Estadística e Informática – INEI elaboró el primer plano estratificado a nivel de ingresos, calculados a partir de los Censos Nacionales de Población y Vivienda del 2007 y la Encuesta Nacional de Hogares 2007.

Dicho trabajo se realizó en base a un modelo econométrico, que consideró 21 indicadores asociados a características demográficas, educación, empleo, equipamiento, vivienda. Los parámetros utilizados fueron construidos en la ENAHO 2007, teniendo el ingreso per cápita por hogar como variable dependiente. Utilizando el modelo econométrico, se estimó el ingreso per cápita promedio por manzana, determinándose así cinco estratos. Para la estratificación de Lima Metropolitana se aplicó la propuesta de Dalenius, cuyo método consiste en definir los límites de cada estrato con la finalidad que la varianza entre las medias de los estratos sea la máxima y la varianza dentro de cada estrato la mínima.

2.2. Mejoras incorporadas

La metodología de la Estratificación de Lima Metropolitana a nivel de manzanas 2013 según los ingresos estimados, incorporó mejoras en cada uno de los procesos; mejoras que redundaron en una mayor precisión y robustez de los estimados y por consiguiente de la estratificación. En primer lugar las fuentes de información por muestreo y las fuentes censales correspondieron al mismo periodo. Por un lado se utilizó el Empadronamiento Distrital de Población y Vivienda 2012-2013, que es una fuente de datos censal recientemente ejecutada en todo el territorio nacional y por otro lado se empleó la Encuesta Nacional de Hogares 2012-2013, potente encuesta de hogares que mide de forma precisa los ingresos per cápita de los hogares. En ambas fuentes de datos no existe un desfase entre los años de ejecución, minimizando así las discrepancias entre las distribuciones de las variables predictivas del modelo de ingreso. El uso de otras fuentes censales y registros administrativos, también correspondientes a los mismos periodos de referencia, permitieron captar las especificidades locales a un nivel de desagregación mínima (conglomerados), mejorando así la capacidad predictiva de los modelos y reduciendo los posibles sesgos.

Otra de las mejoras consistió en utilizar en la etapa de simulación, la información existente de los hogares presentes en la encuesta con el fin de reducir los errores estándar de los estimados (método "empirical best"), mejorando las predicciones de su entorno.

El procedimiento de estimación de ingresos en áreas menores consta de tres etapas básicas. En la primera etapa se examina la similitud de las definiciones y distribuciones de las variables provenientes de la encuesta, comparándolas a la distribución de las mismas variables presentes en el censo. La segunda etapa consiste en el modelamiento de los ingresos mediante regresiones econométricas.

¹ Propuesta por el Banco Mundial y desarrollada por Chris Elbers, Jean Lanjouw y Peter Lanjouw ("Micro-Level Estimation of Poverty and Inequality", Econometrica 2003)

III. METODOLOGÍA

La tercera etapa es la simulación de los ingresos utilizando los parámetros estimados del modelo así como una estimación los diferentes componentes aleatorios. Ello resulta en una estimación de ingresos para cada uno de los hogares y para luego poder ser agregados en los diferentes niveles geográficos. En las diferentes etapas de estimación se hizo uso de las recientes innovaciones en el software Povmap (elaborado por el Banco Mundial) como herramienta fácil de utilizar y eficiente para manipular datos y hacer simulaciones.

Finalmente, la estimación a nivel de manzanas se hizo más robusta porque se homogenizó y armonizó el tamaño de las manzanas, en algunos casos se han agrupado en conglomerados para que la estimación final tenga la robustez requerida y no presente sesgo según las distintas zonas de Lima Metropolitana.

Se implementaron también procesos de validación, con el fin de examinar la validez de los resultados obtenidos en la estimación de los indicadores estratificados a nivel de manzanas según los ingresos estimados.

3.1. Metodología de estimación del ingreso

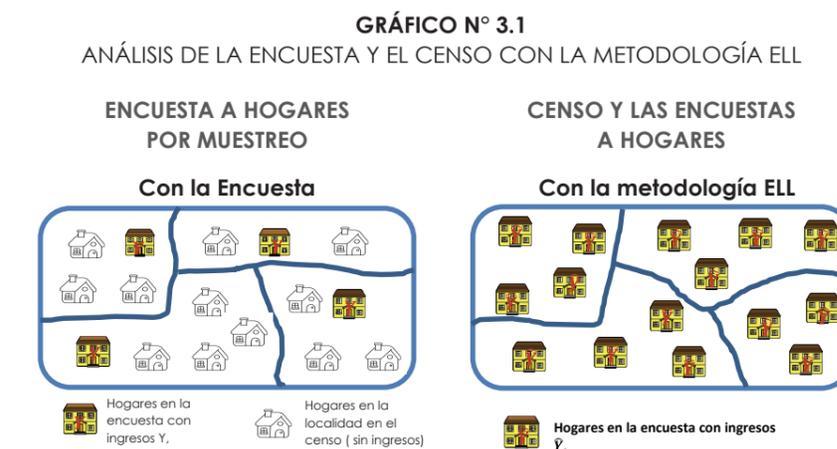
Como se mencionó, los censos de población no captan información sobre los ingresos de los hogares debido a su alta complejidad como para ser medida a través de un censo. Por otro lado, tenemos las encuestas especializadas que si tienen información de ingresos obtenida de forma robusta. Sin embargo, debido al diseño propio de las encuestas por muestreo (como es el caso de la Encuesta Nacional de Hogares), los indicadores de ingresos de los hogares son estimados únicamente con un nivel de representatividad departamental más no distrital y mucho menos en áreas menores que el distrito.

La metodología de estimación de áreas menores (ELL)¹ desarrollada por Chris Elbers, Jean Lanjouw y Peter Lanjouw (Econometrica, 2003), economistas del Banco Mundial dan solución a este problema. La metodología propuesta consiste en combinar información de una encuesta (Encuesta Nacional de Hogares 2012-2013) y de un censo de población (Empadronamiento Distrital de Población y Vivienda 2012-2013). Teniendo en cuenta que en la encuesta de hogares capta el ingreso de los hogares, se estima los modelos predictivos del ingreso per cápita del hogar con información de la encuesta, empleando las variables comunes en el censo, luego aplica los parámetros estimados en los datos censales para predecir el ingreso de cada hogar censado, obteniendo estimaciones de los niveles del ingreso per cápita por hogar promedio a diferentes niveles de desagregación geográfica.

Una de las condiciones de validez de la metodología es que las definiciones de las variables utilizadas en el modelo econométrico tenga definiciones operativas idénticas y que sus distribuciones sean similares (media, varianza en particular). La otra condición necesaria es que la especificación del modelo econométrico así como las variables incluidas en la regresión sean relevantes para la predicción del ingreso. Es necesario por consiguiente que el modelo estimado tenga una capacidad predictiva adecuada (es decir un coeficiente de determinación suficientemente elevado que minimice los errores de predicción), los coeficientes estimados sean significativos y que las variables predictivas sean capaces de captar la heterogeneidad espacial de los niveles de ingresos.

La disponibilidad de bases de diferentes datos censales geo referenciadas y registros administrativos (fuentes detalladas más adelante) ha permitido incluir en el modelo de regresión variables que reflejan las heterogeneidades espaciales al nivel más fino posible (de manzanas y conglomerados). Ello tiene no solamente la ventaja de captar la heterogeneidad local mejorando la capacidad predictiva del modelo econométrico sino también, al estar dichas variables censales exentas de errores muestrales, ello redundará en la mejora de la precisión de las predicciones.

En el siguiente Gráfico N°3.1, se observa un gráficamente la metodología ELL aplicando la encuesta y el censo.



3.2. Fuentes de información

La construcción de la Estratificación de Lima Metropolitana a nivel de manzanas según los ingresos 2013, tiene como fuentes principales al Empadronamiento Distrital de Población y Vivienda 2012-2013 (SISFOH) y la Encuesta Nacional de Hogares 2012-2013, estas dos potentes fuentes de datos fueron homogenizadas a nivel de conglomerados y actualizadas geográficamente.

Se armonizaron estas dos bases con la información proveniente de otros censos y registros administrativos, denominados “fuentes externas” a nivel de conglomerados, centros poblados y distritos. Entre ellos destacan el Censo de Infraestructura Educativa 2013, Censo Escolar 2013, Evaluación Censal de Estudiantes 2012 y 2013, Registro Nacional de Municipales 2014, Censo Nacional a Gobiernos Regionales 2014, Empadronamiento Distrital de Población y Vivienda.

3.3. Elaboración del modelo del ingreso estimado

En la elaboración del modelo predictivo del ingreso de los hogares se procedió en varias etapas. (1) En primer lugar fue necesario seleccionar las variables predictivas que potencialmente correlacionadas con los ingresos de los hogares y que tengan definiciones y distribuciones similares en la ENAHO y el SISFOH, (2) En segundo lugar se determinó el modelo predictivo del ingreso de los hogares y se imputó a cada hogar del SISFOH los coeficientes y los errores estimados a partir de la ENAHO y variables externas provenientes de fuentes censales o administrativas (3) Finalmente, se analizó la sensibilidad de las estimaciones frente a diferentes especificaciones y desagregaciones geográficas.

¹Propuesta por el Banco Mundial y desarrollada por Chris Elbers, Jean Lanjouw y Peter Lanjouw (“Micro-Level Estimation of Poverty and Inequality”, Econometrica 2003)

Asimismo, se examinó la precisión de los indicadores obtenidos. Luego de la fase de estimación e imputación se construyeron los intervalos de ingresos con el fin de clasificar las manzanas según los cinco estratos requeridos.

3.3.1. Selección de variables predictivas del ingreso

Este proceso es uno de los más importantes debido a que exige el conocimiento del marco conceptual del presente estudio. Se prestó especial cuidado en verificar el cumplimiento de las condiciones subyacentes para la validez para la aplicación del método de estimación de áreas menores. También se evaluó las interacciones entre variables predictivas. Hay que destacar que los ingresos del hogar reflejan los recursos monetarios obtenidos por los hogares de sus diferentes fuentes, siendo el ingreso por trabajo la principal. Ello significa que el mapa estratificado refleja de manera directa las fuentes primarias de los ingresos y por ende la sostenibilidad de las condiciones de vida de los hogares.

Igualdad en la definición y distribución de las variables predictivas

A partir de variables existentes con definiciones idénticas en la encuesta y el censo, se construyeron indicadores o variables complejas armonizadas, obteniéndose en un inicio más de mil variables candidatas a predecir el ingreso. Alrededor de 371 variables están relacionadas con las características demográficas del hogar, educación, capital humano, activos físicos y acceso a servicios, características de las viviendas y servicios básicos, empleo e indicadores de capacidad económica. Las fuentes externas censales y administrativas permitieron otros 634 indicadores construidos a partir de las variables que fueron agregadas a nivel de conglomerado y distritos. En la selección de variables predictivas de los ingresos se cuenta con la ventaja de la existencia de números trabajos de estimación empírica siguiendo la tradición de los modelos mincerianos de ingresos.

(i) Características demográficas

Entre las características demográficas y composición de los hogares, se consideraron indicadores correlacionados con los ingresos del hogar como por ejemplo la composición de la población por grupos de edad (niño, adolescente, en edad activa, adulto mayor y sus respectivos ratios). Hogares con mayores tasas de dependencia tienen en promedio menor capacidad a generar ingresos debido a la menor participación en el mercado de trabajo que ello implica. Asimismo, la edad del jefe del hogar, generalmente el principal perceptor de ingresos refleja a la vez su experiencia en el mercado laboral y la variación de los ingresos a lo largo del ciclo de vida. Se consideraron igualmente el tamaño del hogar, la situación de aseguramiento de salud de los miembros del hogar, entre otras.

(ii) Características educativas

Los años de educación tienen una relación directa con los ingresos. Se probaron distintas especificaciones de esta dimensión. Así, considerando el posible efecto de "acreditación" de los niveles de educación en el mercado de trabajo, se categorizó el nivel de educación del jefe del hogar según niveles educativos (primaria, secundaria, superior no universitaria y universitaria). Con el propósito de captar competencias mínimas, se consideró la tasa de analfabetismo. Esta dimensión, que se apoya en la teoría del capital humano formulada por Mincer y Becker, es una de las más estudiadas en la literatura empírica sobre la determinación de los ingresos. Se ha ampliamente demostrado el efecto directo y positivo de la educación, que implica una mayor productividad laboral, sobre los ingresos.

(iii) Características de la Población Económicamente Activa (PEA)

En la medida que en el SISFOH no se cuenta con información de una variable monetaria, se complementaron las variables de las otras dimensiones con variables relacionadas a la condición de la actividad económica de los miembros del hogar. Se considera variables referidas a la rama de actividad de los miembros del hogar: servicios, estado (gobierno), comercial, entre otros. Las diferencias de ingresos entre las actividades económicas en que se emplean los miembros del hogar captan el flujo de ingresos, así como también si son trabajadores dependientes, independientes, desempleados, estudiantes, jubilados.

(iv) Características y servicios básicos de la vivienda

Se construyeron variables relacionadas al tipo de material de construcción predominante en las estructuras de las viviendas (paredes exteriores, techos y pisos), los servicios básicos (agua, desagüe y alumbrado eléctrico), el combustible que utiliza el hogar para preparar sus alimentos, tenencia de la vivienda (casa propia, alquiler, etc.), precariedad de la vivienda en relación a su infraestructura, entre otros. La características de la vivienda están relacionadas por un lado con el ingreso pasado y por otro lado con el acceso a servicios básicos, factor correlacionado al entorno geográfico del hogar, identificando en particular las zonas marginales de la ciudad con poca presencia del estado a través de las empresas de servicio público.

(v) Activos físicos y acceso a servicios

El equipamiento del hogar constituye un reflejo de posibilidades del grado de capital no humano acumulado por el hogar a lo largo de un periodo que va más allá de la coyuntura, reduciendo así el impacto de la volatilidad de los ingresos que caracteriza a la población de menores recursos. Los equipos que posee el hogar, tipos de equipos (celular, televisor a color, equipo de sonido, refrigeradora, computadora, lavadora de ropa), asimismo, los servicios no indispensables del hogar (tv-cable, internet, teléfono fijo) muestran hasta cierto grado el poder de adquisición de los miembros del hogar, debido a que, al hablar de equipamiento no solo se refiere a la adquisición de activos o acceso a servicios sino también al mantenimiento de estos en el tiempo.

(vi) Otros indicadores de la vivienda y hogar

Se tiene indicadores como las Necesidades Básicas Insatisfechas (NBI) y se define como hogares con privación aquellos que tengan al menos una privación en alguno de los cinco componentes de las NBI (precariedad de las viviendas, el hacinamiento, la falta de servicios higiénicos adecuados, la no asistencia escolar de los niños, la dependencia económica) y en situación de miseria los hogares que tengan dos o más de los indicadores simples de necesidades básicas insatisfechas. Estos índices de NBI permiten jerarquizar los espacios geográficos con mayor pobreza estructural en relación a las carencias.

(vii) Otras variables

Se utilizó otras fuentes censales y registros administrativos, también correspondientes a los mismos periodos de referencia, que han permitido captar las especificidades locales a un nivel de desagregación mínima (conglomerados), mejorando así la capacidad predictiva de los modelos y precisión de las estimaciones al no comportar un error muestral. Entre las fuentes externas se tiene el Censo de Infraestructura Educativa 2013, Censo Escolar 2013, Evaluación Censal de Estudiantes 2012 y 2013, Registro Nacional de Municipales 2014, Censo Nacional a Gobiernos Regionales 2014, Empadronamiento Distrital de Población y Vivienda 2012-2013, las variables fueron armonizadas en la encuesta y en el censo.

Entre las variables de fuentes externas agregadas a nivel de conglomerado y distrito, se tienen variables altamente correlacionadas con la variable ingresos. Entre ellos el ratio del número de alumnos entre el número de aulas, el tiempo de recorrido del local escolar a la capital del distrito, características de docencia, características de las instituciones educativas, número de matriculados, número de docentes, bienes del local escolar, evaluación de comprensión lectora y matemática; variables de competencias y funciones de la municipalidad, gestión y políticas de desarrollo, sistema informático y comunicación de la municipalidad, variables de información general del gobierno municipal, desarrollo económico, servicios sociales y municipales y variables de características de la vivienda, hogar y población a nivel de conglomerado y distrito, adicional a ello se cuenta con variables de estrato socioeconómico.

Verificación de la igualdad de la distribución de las variables predictivas

Luego de definir las posibles variables predictivas del ingreso estimado, es necesario realizar el primer filtro de selección de variables comunes en la encuesta y el censo, considerando la misma definición de las variables y tener distribuciones semejantes. Se verificó que el valor promedio de las variables en el SISFOH, se encontraba dentro de los intervalos de confianza (al 95%) del valor obtenido en la ENAHO.



En el caso de las variables continuas, se efectuaron los test de medias (t-student) teniendo como hipótesis nula que no existan diferencias significativas al 95% de confianza y se evaluó los gráficos de Kernel de cada variable continua comparable en la encuesta y el censo; mientras que, en el caso de las variables cualitativas se examinaron los resultados de los test de chi cuadrado.

Como primer resultado de evaluación con los estadísticos, se obtiene 68 variables sobre un total de 371 variables internas iniciales comparables en la ENAHO y SISFOH que lograron pasar el test de igualdad de medias, distribución de Kernel y la chi cuadrado.

Selección de variables en el modelo predictivo

Considerando que para cada uno de las dimensiones se contaba con varias maneras de calcular el mismo indicador captando el mismo fenómeno, era necesario proceder a una selección de las mismas con el fin de evitar los problemas derivados de la colinealidad existente entre ellas. En efecto, la presencia de variables altamente colineales provocan inestabilidad en los parámetros estimados Para salvar este problema se empleó las técnica de selección de variables "stepwise".

Una regla empírica, citada por Kleinbaum², consiste en considerar que pueden existir problemas de colinealidad si algún factor de Inflación de varianza es superior a 10. Cumpliendo con la regla de multicolinealidad en la regresión se filtraron 340 variables, es decir, pasaron el proceso de limpieza de variables altamente colineales.

3.3.2. Proceso de selección del modelo del ingreso e imputación en el SISFOH

Para las estimaciones de los ingresos es necesario realizar un modelo econométrico que prediga el ingreso per cápita del hogar y seleccione óptimamente las variables que permitan obtener buenos resultados en pruebas de bondad de ajuste. La técnica de estimación stepwise³ maximiza el R2 ajustado, conservando en la regresión final únicamente las variables que son individualmente significativas estadísticamente. Este método de selección conviene en los casos en que se trata de estimar un modelo puramente predictivo. El riesgo, que ha sido evaluado, es que el ajuste sea elevado únicamente en la muestra considerada y que, transpuesta a otra muestra, los R2 ya no sean elevados ni la selección de variables la más óptima. Para ello se implementó una comparación de la capacidad del modelo predictivo dentro respecto a la predicción fuera de la muestra (ver sección 4 en la que se presenta el análisis de sensibilidad).

Modelo estadístico y bondad de ajuste

Para este documento se desarrolló el modelo de regresión de errores anidados propuesto por Elbers, Lanjouw y Lanjouw (2003)⁴. Como variable a predecir se tiene el logaritmo del ingreso per cápita del hogar para así atenuar el peso de los valores extremos en la regresión y por ende en los coeficientes.

Dentro del proceso de imputación se aplicó el método "Empirical Best", denominado así como el mejor estimador desarrollado por Molina y Rao, (2010). Asumiendo que los errores tienen una distribución según la ley normal, se minimizaron los errores de estimación de áreas menores mejorando las predicciones de su entorno al utilizarse la información de ingresos recogida directamente a través de la muestra de los hogares presentes en la encuesta.

En el Cuadro N°3.1 se observa el resultado del modelo econométrico final en la encuesta, las variables son significativas individualmente y en su conjunto. El número de observaciones en la encuesta es de 5 949 hogares para Lima Metropolitana, considerado número suficiente de casos para una estimación robusta de los parámetros de regresión. La bondad de ajuste del modelo predictivo medido a través del coeficiente de determinación ajustado es de 0,55; es decir más de la mitad de la varianza de los ingresos es explicada por el modelo. Dicho coeficiente, dada la experiencia internacional, es considerado adecuado por estar dentro de los parámetros establecidos. El número de variables en el modelo predictivo son 27 variables, el error cuadrático medio es 0,25.

CUADRO N°3.1
MODELO DEL LOGARITMO DEL INGRESO

Ind. Nro.	Variables explicativas	Coefficiente	Std. Err.	t	Prob >t
1	INTERCEPTO	5,334	0,20	26,8	0,00
2	LTAMHOG	-0,613	0,02	-30,5	0,00
3	SEGUROJEFE	0,231	0,02	10,8	0,00
4	SEGUROJEFE_MBB	-0,140	0,03	-4,8	0,00
5	CABLE	0,203	0,02	9,7	0,00
6	CABLE_MBB	-0,057	0,03	-1,9	0,05
7	TVCOLOR	0,096	0,03	2,8	0,01
8	IEQUIPO_MBB	0,089	0,01	10,7	0,00
9	TECHO1_MBB	-0,142	0,03	-4,1	0,00
10	TECHO_PREC	-0,138	0,03	-4,7	0,00
11	TENPROP	0,104	0,01	7,4	0,00
12	VIVIENI	0,159	0,02	8,6	0,00
13	ABATALUMGAS	0,046	0,02	2,4	0,02
14	PARED_PREC	-0,067	0,02	-2,9	0,00
15	ELECTRI2_MBB	-0,313	0,05	-6,7	0,00
16	EDUPRIM3	0,060	0,01	7,0	0,00
17	EDUSEC1	0,050	0,01	6,0	0,00
18	EDUSUP2_OCUC	0,202	0,01	18,6	0,00
19	TASAREMUN_MBB	0,692	0,04	17,5	0,00
20	SIS_COMERCI	0,402	0,17	2,3	0,02
21	SIS_PCOMPUTAD	0,529	0,06	9,6	0,00
22	SIS_PEDUDEF1599	0,007	0,00	2,2	0,03
23	SIS_RAMAS	3,012	0,43	7,0	0,00
24	SIS_RAMAS4	6,517	0,88	7,4	0,00
25	ECE_MTNESTPR_2012	-0,003	0,00	-4,1	0,00
26	ECE_MTNESTSAT_2013	0,002	0,00	2,7	0,01
27	CENG_C9P1_26	-0,343	0,08	-4,3	0,00
28	REN_BIB_MUN	0,053	0,02	2,3	0,02

- 1 Logaritmo tamaño del hogar
- 2 Jefe del hogar tiene seguro de salud
- 3 Jefe del hogar tiene seguro de salud en el estrato medio bajo y bajo
- 4 Tiene tv - cable
- 5 Tiene tv - cable en el estrato medio bajo y bajo
- 6 Tiene televisor a color
- 7 Número de equipos que tiene el hogar en el estrato medio bajo y bajo
- 8 Techo de concreto armada en el estrato medio bajo y bajo
- 9 Precariedad del techo (madera, tejas, planchas de calamina, caña o estera, entre otros)
- 10 Vivienda propia totalmente pagada
- 11 Vivienda de material noble
- 12 Abastecimiento de agua y servicios higiénicos por red pública dentro de la vivienda, alumbrado eléctrico y cocina con gas

³Thompson, B. (2001). Significance, effect sizes, stepwise methods, and other issues: Strong arguments move the field. *Journal of Experimental Education*.

⁴Elbers, Chris, Peter Lanjouw, Johan Mistiaen, Berk Özler y Kenneth Simler, Are Neighbors Equal? Estimating Local Inequality in Three Developing Countries, International Food Policy Research Institute, FCND DP No.147, abril de 2003.

²D.G. Kleinbaum, L.L. Kupper, K.E. Muller. *Applied Regression Analysis and Other Multivariable Methods*. PWS-KENT Publishing Company. 1988.

- 13 Precariedad de la pared (piedra o sillar con cal, adobe o tapia, quincha, piedra con barro, madera, estera, entre otros)
- 14 El tipo de alumbrado es eléctrico en el estrato medio bajo y bajo
- 15 Número de miembros del hogar entre 15 y 64 años con primaria completa
- 16 Número de miembros del hogar en educación secundaria
- 17 Número de miembros del hogar de 18 a más años de edad con superior universitaria completa y ocupados
- 18 Tasa de ocupados que perciben ingresos entre el total de miembros del hogar en el estrato medio bajo y bajo
- 19 Proporción de miembros del hogar entre 15 y 99 años que trabajan en el sector comercio en el conglomerado
- 20 Proporción de hogares que tienen computadora en el conglomerado
- 21 Proporción de hogares con déficit educativo de la población de 15 a más años de edad en el conglomerado
- 22 Proporción de miembros hogar por distrito que trabajan en el sector de servicios en el conglomerado
- 23 Proporción de miembros hogar por distrito que trabajan en otros sectores económicos en el conglomerado
- 24 Calificación en proceso del total de alumnos en matemática
- 25 Calificación satisfactorio del total de alumnos en matemática
- 26 La municipalidad otorgó licencias para establecimiento de cabinas públicas
- 27 Municipalidades que tienen Biblioteca Municipal

Imputación (bootstrap)

Una vez estimado el modelo predictivo a partir de las variables internas (ENAHO) y externas (datos censales y administrativos), la etapa siguiente consiste en estimar los ingresos en cada hogar del censo. Para ello se imputan los coeficientes de la regresión a las variables comunes. Es necesario igualmente imputar una estimación de los errores.

La imputación de los errores se realizó mediante un proceso de simulación por "bootstrap". La simulación consiste en generar valores de los parámetros a partir de réplicas aleatorias de la base censal a la se imputaran las estimaciones. Así, se replicó 100 veces las estimaciones de suerte que se pudo estimar los ingresos promedio y las desviaciones estándar en los niveles geográficos deseados.

En el proceso de imputación de ingresos se excluyeron los valores extremos predichos por el modelo, lo cuales fueron identificados comparando los valores predichos con los rangos máximo y mínimo de los ingresos obtenidos por la ENAHO. Con ello se recortaron los valores extremos estimados (procedimiento llamado "trimming" en el software Povmap).

3.4. Evaluación de manzanas censales

Con la estimación del modelo en la ENAHO e imputación de los parámetros en cada hogar del SISFOH se obtiene el ingreso estimado per cápita a nivel del hogar. Con ello ya se podría promediar el ingreso del hogar a nivel de manzanas censales, pero antes de ello es necesario evaluar la distribución de viviendas por manzana censal para tener estimaciones más robustas a ese nivel.

En Lima Metropolitana se tiene más de 90 mil manzanas censales, en promedio se tiene 25 viviendas por manzanas, la mediana es 16 y la moda 6. El número de viviendas por manzana es de una vivienda a 2 421 viviendas con ocupantes presentes.

Después de una evaluación del número de viviendas por manzana para cada distrito de Lima Metropolitana se obtuvieron los siguientes resultados:

- En 20 distritos se encontró que, en promedio, más del 60% de manzanas tienen entre uno a 10 hogares. Por ello, para una mayor precisión, en estos casos el cálculo del ingreso promedio per cápita del hogar se efectuó a nivel de conglomerados.

- En 16 distritos con manzanas censales que tienen entre uno a 10 hogares, más del 80% de las manzanas son similares en su estrato. La mediana del número de hogares por manzana es de 25 hogares, por lo cual, aquellas manzanas con menos de 10 hogares se agrupan con manzanas de su entorno (cercanía espacial) perteneciente a su mismo estrato.

- Para 14 distritos se realizó un agrupamiento manual por cercanía espacial de manzanas (pequeñas) que tenían entre uno a 10 hogares.

En resumen de las 36 744 manzanas de Lima Metropolitana tienen entre uno a 10 hogares, se realizó el agrupamiento espacial de manzanas para el 4,8%. El 77,8% de manzanas fueron agrupadas según el conglomerado al cual pertenecen y el 16,1% se le asignó por cercanía espacial el estrato de su entorno.

Para cada manzana censal se obtuvo el ingreso promedio per cápita del hogar calculado a nivel de manzanas, manzanas agrupadas y conglomerados según la condición de la manzana en los distritos mencionados anteriormente. Con la homogenización y armonización del tamaño de las manzanas la estimación final es más robusta y reduce el sesgo según las distintas zonas de Lima Metropolitana.

3.5. Estratificación del ingreso

En la medida que, por un lado se requiere robustez en la identificación de los hogares según sus niveles de pobreza y, por otro lado se ha definido una focalización que distinga 5 estratos poblacionales, se empleó el método de Dalenius-Hodges (1959) en la constitución de los rangos que definen los cinco estratos de ingresos.

El método consiste en la formación de estratos de manera que la varianza de las medias sea mínima al interior del estrato y máxima entre cada uno de ellos, es decir, formar estratos lo más homogéneos posible.

En el Cuadro N° 3.2, utilizando la propuesta de Dalenius se construye cinco estratos de ingresos a nivel de manzanas censales, el estrato alto de 2 192,20 soles a más soles, el estrato medio alto de 1 330,10 soles hasta 2 192,19 soles, el estrato medio de 899,00 soles hasta 1 330,09 soles, el medio bajo de 575,70 soles hasta 898,99 soles y el bajo hasta 575,69 soles a precios de Lima Metropolitana.

CUADRO N°3.2
ESTRATIFICACIÓN DE INGRESOS POR EL MÉTODO DE DALENIUS

ESTRATO	RANGO DE INGRESOS PER CÁPITA (Soles)		
ALTO	2 192,20	a	más
MEDIO ALTO	1 330,10	-	2 192,19
MEDIO	899,00	-	1 330,09
MEDIO BAJO	575,70	-	898,99
BAJO	Hasta		575,69

En el Cuadro N° 3.3, a nivel de manzanas el estrato más alto contiene a 4,7%, el medio alto a 9,1%, el medio a 19,5%, el medio bajo a 36,9% y el bajo a 29,9% de manzanas censales. Si consideramos la estratificación a nivel de personas se tiene que en el estrato alto representa a 5,2%, el medio alto a 14,1%, el medio a 31,1%, medio bajo a 36,4% y el bajo a 13,3% de los miembros del hogar.

CUADRO N°3.3
ESTRATIFICACIÓN DE INGRESOS POR EL MÉTODO DE DALENIUS

ESTRATO	PERSONAS		HOGARES		MANZANAS	
	Absoluto	%	Absoluto	%	Absoluto	%
ALTO	381 423	5,2	123 165	6,1	4 250	4,7
MEDIO ALTO	1 040 030	14,1	305 951	15,3	8 207	9,1
MEDIO	2 298 076	31,1	624 359	31,2	17 642	19,5
MEDIO BAJO	2 687 240	36,4	685 499	34,2	33 460	36,9
BAJO	982 231	13,3	264 773	13,2	27 041	29,8
TOTAL	7 389 000	100,0	2 003 747	100,0	90 600	100,0

IV. ANÁLISIS DE PRECISIÓN Y ROBUSTEZ

Con el fin de evaluar la robustez de las estimaciones se hicieron varios test de sensibilidad y se evaluó la precisión y plausibilidad de las estimaciones. En primer lugar se compararon los valores predichos por la regresión con los valores observados en la misma encuesta ENAHO. Enseguida se evaluó en qué medida el método "stepwise" de selección de variable permitía o no una imputación de los coeficientes y errores estimados en un conjunto distinto de observaciones pero con similar distribución. Finalmente, se compararon las distribuciones de las estimaciones obtenidas mediante la regresión con la distribución obtenida en la imputación en áreas menores.

4.1. Análisis de sensibilidad

Como se vio en el capítulo anterior, se realizó la selección de variables por el método "stepwise", que permite optimizar la combinación de variables para tener el mejor R2 ajustado posible. Se debía sin embargo descartar sin embargo que el ajuste del modelo sea demasiado específico a los datos utilizados en la estimación y por lo tanto la predicción podría ser no muy buena con otro conjunto de observaciones.

- Para la selección de una sub muestra y probar el modelo econométrico sobre una base de datos que tiene la misma característica de distribución, se divide la muestra de la ENAHO 2012-2013 en dos partes iguales obtenidas de forma aleatoria.
- Enseguida se estimaron los coeficientes de la regresión en base únicamente de la primera muestra y luego se imputaron los coeficientes y errores estimados a la segunda mitad de la muestra.
- Finalmente, se compararon los valores predichos en esta segunda submuestra con los valores observados en la ENAHO.

En el Cuadro N° 4.1, se presentan los ingresos estimados y observados de la sub muestra de la base 2 donde se imputo los coeficientes estimados de la sub muestra de la base 1. Los test de diferencia de los ingresos estimados fuera de la muestra respecto a los observados en la ENAHO muestran que no existen diferencias significativas entre las medias de los ingresos en ambos caso por lo cual concluyó que el método de stepwise no indujo un sobre-ajuste de la regresión y que los coeficientes y errores estimados podían ser imputados a otro conjunto de observaciones con distribución semejante.

CUADRO N°4.1
COMPARACIÓN DEL INGRESO PROMEDIO ESTIMADA Y OBSERVADA EN LA SUB MUESTRA DE LA ENAHO

FUENTE	Ingreso Promedio	Intervalos de confianza al 95%		Variable Porcentual	n_ P >t Significancia
		Límite inferior	Límite superior		
Observado base 2	1 230	1 161	1 299	-2,2	0,5
Estimado base 2	1 203	1 154	1 253		

* Diferencia significativa (p < 0.10).
** Diferencia altamente significativa (p < 0.05).
*** Diferencia muy altamente significativa (p < 0.01).
Fuente: Instituto Nacional de Estadística e Informática.

Asimismo, se comparó el ingreso promedio de la ENAHO observada y estimada del total de la muestra. En el Cuadro N°4.2, se observa que no existe diferencia significativa entre las dos medias del ingreso.

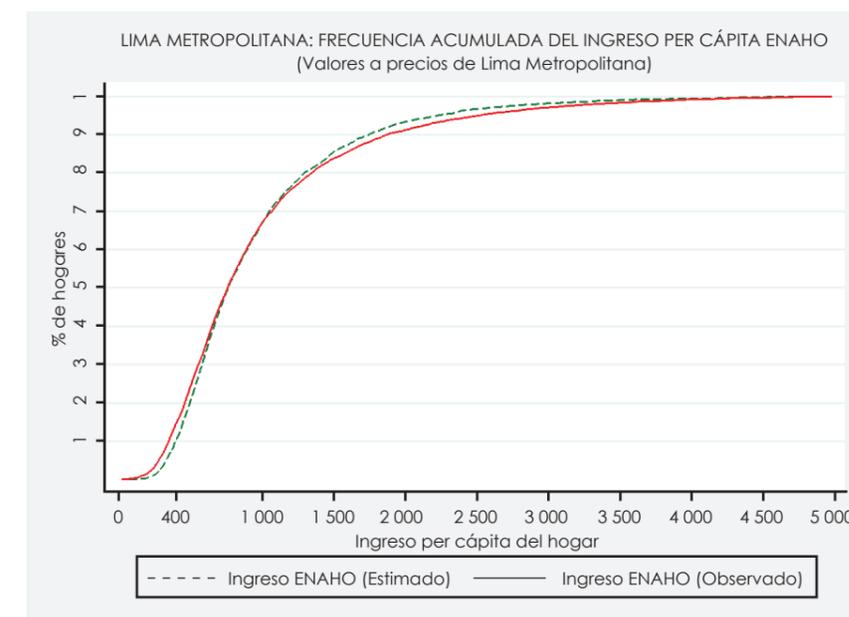
CUADRO N°4.2
COMPARACIÓN DEL INGRESO PROMEDIO DE LA ENAHO OBSERVADA Y ESTIMADA

FUENTE	Ingreso		Intervalos de confianza al 95%		P >t Significancia
	Promedio	Coefficiente variable	Límite inferior	Límite superior	
ENAHO estimada	1 072,66	1,7	1 037,2	1 108,1	0,87
SISFOH estimada	1 069,10	1,1	1 045,2	1 093,0	

* Diferencia significativa (p < 0.10).
** Diferencia altamente significativa (p < 0.05).
*** Diferencia muy altamente significativa (p < 0.01).
Fuente: Instituto Nacional de Estadística e Informática.

En el siguiente GRÁFICO N°4.1 se tiene la frecuencia acumulada del ingreso promedio per cápita del hogar en la muestra ENAHO observada y estimada. Se puede constatar que la curva de frecuencia acumulada de ambas distribuciones coinciden fuertemente.

GRÁFICO N°4.1
FRECUENCIA ACUMULADA DEL INGRESO PER CÁPITA DEL HOGAR ESTIMADO Y OBSERVADO EN LA ENAHO



Finalmente, se comparó el ingreso promedio de la ENAHO estimada y el ingreso promedio estimado en el SISFOH. En el Cuadro N° 4.3, se aprecia que no existe diferencia significativa entre las dos medias del ingreso.

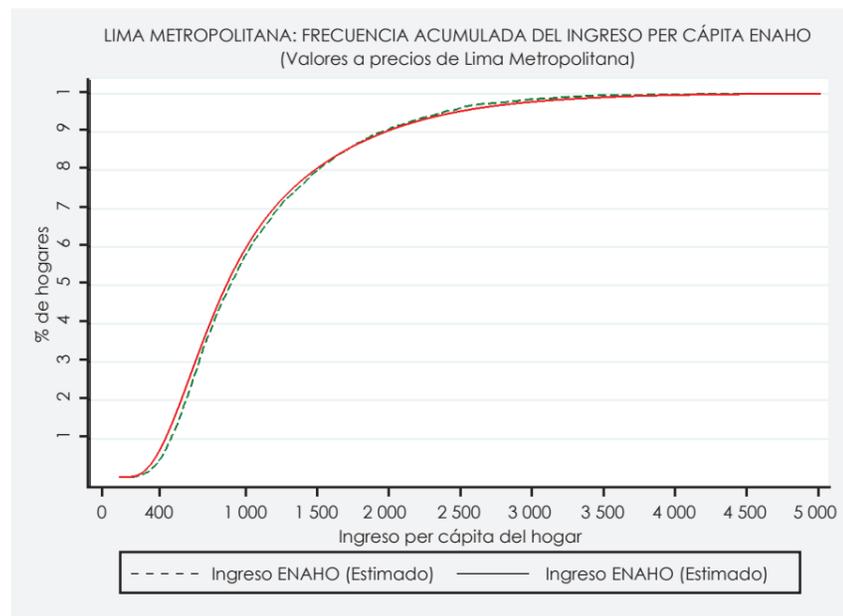
CUADRO N°4.3
COMPARACIÓN DEL INGRESO PROMEDIO DE LA ENAHO ESTIMADA Y SISFOH ESTIMADA

FUENTE	Ingreso		Intervalos de confianza al 95%		P >t Significancia
	Promedio	Coefficiente variable	Límite inferior	Límite superior	
ENAHO observada	1 059,88	2,2	1 014,9	1 104,9	0,664
SISFOH estimada	1 072,66	1,7	1 037,2	1 037,2	

* Diferencia significativa (p < 0.10).
** Diferencia altamente significativa (p < 0.05).
*** Diferencia muy altamente significativa (p < 0.01).
Fuente: Instituto Nacional de Estadística e Informática.

En el siguiente Gráfico N°4.2, se tiene la frecuencia acumulada del ingreso promedio per cápita del hogar estimado en la muestra ENAHO y el ingreso promedio per cápita del hogar estimado en el SISFOH.

GRÁFICO N°4.2
FRECUENCIA ACUMULADA DEL INGRESO PER CÁPITA DEL HOGAR ESTIMADO EN LA ENAHO Y EL SISFOH



V. ANÁLISIS DE LOS RESULTADOS

5.1. Comparación estratificación del ingreso y gasto 2013

Una primera verificación de la plausibilidad de los resultados obtenidos consistió en comparar la clasificación de los ingresos promedios a nivel de manzanas en los distintos estratos según nuestras estimaciones por ingresos con aquella que resulta utilizando los gastos per cápita de los hogares para el mismo año 2013. En el cuadro N°5.1 se constata que el 70,4% (63 804 manzanas) tienen la misma estratificación según el ingreso per cápita y según el gasto, es decir, menos de un tercio (29,1%) de manzanas pertenecen a estratos diferentes. Ello confirma que existe una relación significativa entre la estratificación del ingreso y el gasto (V de Cramer=0,68),

pero esta relación no es perfecta, es decir a medida que aumenta la escala del nivel socioeconómico del ingreso aumenta la escala de nivel socioeconómico del gasto.

Así mismo se evaluó la relación entre los valores del ingreso y el gasto al interior de los estratos (mayor a 0,4) y a nivel global (0,9) observándose que existe una fuerte relación lineal significativa. Cabe señalar que el rango de variación de los gastos es menor al de los ingresos, dada la naturaleza de los mismos.

CUADRO N°5.1
COMPARACIÓN DE LA ESTRATIFICACIÓN DEL INGRESO Y EL GASTO ESTIMADO EN EL SISFOH 2013

Estrato	Estratificación 2013 - ingreso					Total	
	1	2	3	4	5		
Estratificación del Gasto	1	3 592	1 072	5	2	4 671	
	2	645	6 652	4 322	152	2	11 773
	3	13	420	9 876	4 504	105	14 918
	4		60	3 379	20 746	3 996	28 181
	5		3	60	8 056	22 938	31 057
Total	4 250	8 207	17 642	33 460	27 041	90 600	

5.2. Comparación de la estratificación de los ingresos estimados 2013 y la estratificación multivariada 2007

Como ejercicio complementario, se compararon los resultados de la estratificación multivariada 2007 y 2013. Se constata que el 61,5% de manzanas se mantuvo en mismo estrato, evidenciando que existe una relación entre estas 2 estratificaciones (V de Cramer = 0,59), la diferencia se debe que la estratificación multivariada 2007 se obtuvo a nivel de conglomerado, donde el 36,2% de manzanas cambiaron en una posición en su estrato.

CUADRO N°5.2
COMPARACIÓN DE LA ESTRATIFICACIÓN DE LOS INGRESOS ESTIMADOS 2013 Y LA ESTRATIFICACIÓN MULTIVARIADA 2007

Estrato	Estratificación 2013 - ingreso					Total	
	1	2	3	4	5		
Estrato Socioeconómico multivariado 2007	1	3 775	2 675	254	8	6 712	
	2	413	4 205	2 838	273	78	7 807
	3	28	1 042	9 008	6 051	387	16 516
	4	25	198	4 846	15 475	3 281	23 825
	5	9	87	696	11 653	23 295	35 740
Total	4 250	8 207	17 642	33 460	27 041	90 600	

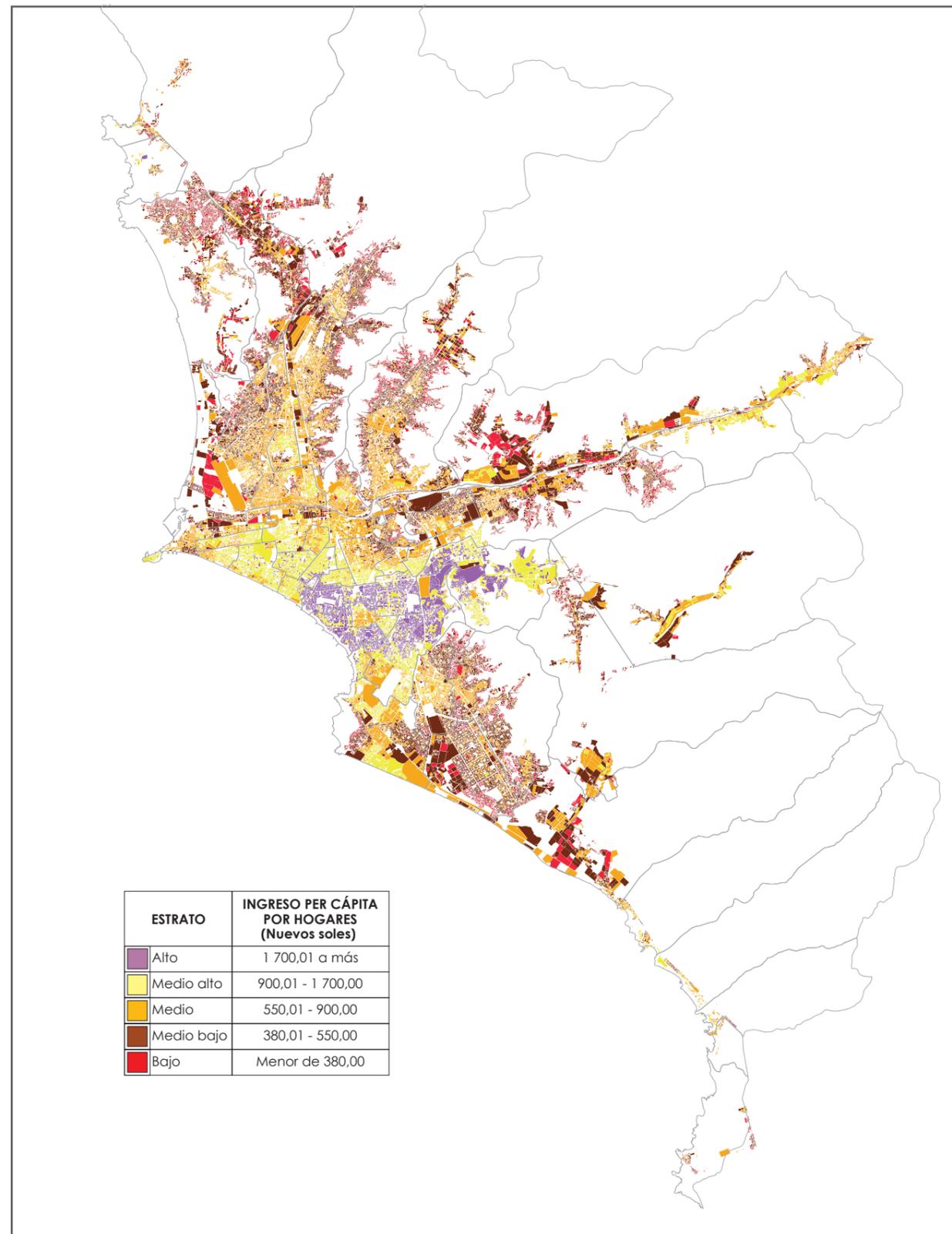
5.3. Comparación de la estratificación de los ingresos estimados del 2007 y 2013

Finalmente se compararon las distribuciones de manzanas según la estratificación por niveles socioeconómicos (estratificación multivariada) utilizada en el mapa estratificado en base al censo del 2007 y la estratificación según ingresos del presente documento. Aunque no existe razones para esperar una fuerte correlación debido al importante crecimiento de los ingresos durante este periodo y al hecho que los criterios de estratificación no son los mismos, se constató que el 58,6% de manzanas coinciden, lo que confirma que existe relación significativa entre la estratificación 2007 y 2013 (V de Cramer =0,56). Mientras que en el Mapa de Estratificación de Lima Metropolitana 2013, se observa más a detalle que en la periferia de Lima se concentran los estratos más bajos, esto se debe principalmente en aquellos distritos donde la pobreza es más alta. Entre los distritos donde se concentra los estratos más bajos tenemos en la zona Norte: Ancón, Ventanilla, Puente Piedra, San Martín de Porres, Carabayllo y Comas, en el Este: San Juan de Lurigancho, Ate y Lurigancho, en la zona Sur: Pachacamac, Villa María del Triunfo, San Juan de Miraflores y Villa el Salvador

CUADRO N°5.3
COMPARACIÓN DE LA ESTRATIFICACIÓN DE LOS INGRESOS ESTIMADOS DEL 2007 Y 2013

Estrato	Estratificación 2013 - ingreso					Total	
	1	2	3	4	5		
Estratificación 2007	1	2 446	560	34	16	6	3 062
	2	1 642	5 174	1 347	199	74	8 436
	3	51	1 783	10 294	5 466	1 127	18 721
	4	7	144	4 153	14 822	5 565	24 691
	5	2	48	359	5 787	7 399	13 595
Total	4 148	7 709	16 187	26 290	14 171	68 505	

**ESTRATIFICACIÓN A NIVEL DE MANZANA
POR INGRESO PER CÁPITA DEL HOGAR, 2007**



**ESTRATIFICACIÓN A NIVEL DE MANZANA
POR INGRESO PER CÁPITA DEL HOGAR, 2013**

