

# Metodología para la estimación de la tasa de capitalización

Unidad Administrativa Especial de Catastro Distrital

Grupo estadístico - Observatorio técnico catastral

01 de Junio de 2021

## Índice

<b>1. Introducción</b>	<b>1</b>
<b>2. Objetivos</b>	<b>3</b>
2.1. Objetivo general . . . . .	3
2.2. Objetivos específicos . . . . .	3
<b>3. Revisión de la base de ofertas</b>	<b>4</b>
3.1. Controles de calidad sobre la base . . . . .	4
<b>4. Modelo aplicado</b>	<b>8</b>
4.1. Apartamentos . . . . .	10
4.2. Casas . . . . .	18
<b>5. Conclusiones</b>	<b>28</b>
<b>6. Trabajo futuro</b>	<b>31</b>
<b>7. Anexo 1</b>	<b>31</b>
7.1. Modelo lineal mixto . . . . .	31

## Resumen

Your abstract.

## 1. Introducción

En la Unidad Administrativa Especial de Catastro Distrital (UAECD) se cuenta con información de ofertas de fuentes secundarias y de fuentes primarias. Las fuentes primarias corresponden a información recolectada en campo por cuenta de la UAECD, mientras que

las secundarias corresponden a información que se recolecta a partir de otras fuentes con las cuales se establecen convenios.

Para diferentes procesos de la Unidad se hace uso de la información recolectada con respecto a ofertas de venta de predios en la ciudad, como por ejemplo en la estimación de parámetros de algunos modelos econométricos que estiman el valor integral de predios en propiedad horizontal. Sin embargo dentro de estos procesos no se tiene en cuenta la información recolectada de ofertas de arriendo, por lo cual este documento tiene como objetivo estimar la tasa de renta inmobiliaria que establezca una relación entre los valores de arriendo y los valores comerciales de los predios en la ciudad de Bogotá, de tal forma que se pueda utilizar la información de ofertas de arriendo en procesos donde se requieren valores comerciales totales de los predios.

Denotando como  $Z_i$  y  $X_i$  al valor comercial y el valor de arriendo del  $i$ -ésimo predio, respectivamente, se busca obtener un coeficiente  $\beta$ , tal que se pueda obtener una ecuación que relacione a  $Z_i$  y  $X_i$  como se muestra en la ecuación (1). De esta manera, de una forma natural se puede pensar en el uso de los rudimientos correspondientes a las metodologías de modelos lineales (Ver Melo et al. [2007]) que permitan estimar el coeficiente  $\beta$ . Este ejercicio se debe realizar para diferentes desagregaciones de predios en la ciudad, tales como usos, estratos y/o desagregaciones geográficas como lo son las localidades y sectores catastrales.

$$Z_i = \beta X_i \quad (1)$$

En la información disponible para la realización de este documento, se encuentran las ofertas recolectadas por la UAECB tanto de fuentes primarias como secundarias (Finca Raíz, Galería Inmobiliaria y Properati) desde el primero de enero de 2017 hasta el 31 de diciembre de 2020. Una limitante que se debe tener en cuenta es que los predios no tienen información de arriendo y venta de manera simultánea, lo que quiere decir que para cada oferta solamente se tiene una de las dos mediciones. En este documento únicamente se tendrán en cuenta las ofertas correspondientes a predios residenciales, es decir apartamentos y casas.

Dentro de la literatura se encuentran diversos ejemplos de datos donde se hacen mediciones o encuestas a grupos de individuos periódicamente, pero no necesariamente a las mismas personas, lo cual imposibilita hacer seguimientos sobre unidades en particular, de manera similar a la problemática presentada, donde no se cuenta con información de arriendo y venta en los mismos predios. Deaton [1985] propone un modelo para una muestra “pseudo-panel”, que se construye con lo que denomina como cohortes, las cuales representan grupos de individuos con características similares, con la condición que cada individuo pertenezca únicamente a una cohorte durante todo el análisis y mediante estos grupos se ajustan efectos dentro de la regresión correspondientes a cada una de las diferentes cohortes. Muchas publicaciones que abordan diferentes contextos se han elaborado a partir del modelo propuesto por Deaton [1985] (Ver Tsai et al. [2014], Tovar et al. [2012], Sprietsma [2012], Antman and McKenzie [2007], entre otros). La propuesta utilizada en este documento se basa en el uso de estas cohortes, agregando la información a nivel de sector catastral, tomando el promedio o la mediana como medidas de resumen, tanto para valores de arriendo como para valores de venta.

Las metodologías de regresión ajustadas por mínimos cuadrados ordinarios tienen como supuestos la independencia entre mediciones, lo cual no se cumple en este caso, ya que un sector catastral puede tener información en las diferentes vigencias, por lo que se propone la inclusión de factores en la regresión que permitan captar el efecto de las medidas repetidas sobre la variable respuesta. De esta manera, la metodología a utilizar corresponde a los modelos de efectos mixtos, donde los efectos fijos van a ser el conjunto de factores asociados al estrato, localidad y el valor de arriendo, mientras que los efectos aleatorios corresponden a la relación entre el valor de arriendo y de venta, segmentada para cada uno de los sectores catastrales.

La programación realizada se realizó en el software R [R Core Team, 2021], en donde el procesamiento de información se llevo a cabo mediante los paquetes relacionados en el *tidyverse* [Wickham et al., 2019], mientras que los controles de calidad utilizaron funciones de la librería *qcc* y *RecordLinkage* (Ver Scrucca [2004] y Sariyar and Borg [2020]). Por último, la construcción del modelo se realizó mediante el paquete *nlme* Pinheiro et al. [2021].

El presente documento contiene en la sección la introducción, seguido de la sección donde se presentan los objetivos del estudio. En la sección 7 se hace una revisión metodológica de las metodologías estadísticas a utilizar. Por otro lado, la sección

## 2. Objetivos

En esta sección se presentan los objetivos a tener en cuenta dentro del estudio, tanto de forma general, como de manera desagregada en los objetivos específicos.

### 2.1. Objetivo general

Estimar un índice que permita relacionar los valores de arriendo de un inmueble con su valor comercial para las diferentes zonas, usos y estratos de la ciudad.

### 2.2. Objetivos específicos

1. Hacer una revisión de la literatura para obtener un estado del arte, con el objetivo de tener en cuenta posibles metodologías que puedan ayudar a completar los otros objetivos específicos y/o puedan servir como metodologías de referencia a la hora de realizar una comparación dentro del mismo estudio.
2. Realizar un análisis descriptivo de la base de datos disponible para la realización del ejercicio.
3. Realizar una validación de las ofertas contenidas dentro de la base de datos, con el objetivo de detectar aquellas ofertas atípicas que puedan sesgar los resultados del estudio.
4. Depurar la base de datos compartida por el OTC.

### 3.1 Controles de calidad sobre la base

5. Realizar el cálculo de las tasas de renta para diferentes desagregaciones de los predios en la ciudad (Uso-Estrato-Localidad/Sector).
6. Presentación y documentación metodológica de resultados.
7. Evaluar la consistencia de los resultados para calcular cifras en vigencias futuras.

## 3. Revisión de la base de ofertas

La base de datos consta originalmente con 1.390.326 registros, incluyendo todos los tipos de predios. Un primer filtro realizado corresponde a dejar aquellas ofertas de predios residenciales, es decir, cuyo tipo de inmueble es apartamento o casa. El total de registros residenciales es 1.072.252. Por otro lado, se tuvo en cuenta diferentes exclusiones para eliminar inconsistencias, dentro de las cuales se incluyen los casos que no tienen estrato, sin área construida, sin identificador del lote (barmanpre) y ofertas rurales. Es importante tener en cuenta que las exclusiones no son excluyentes, lo que quiere decir que un predio puede tener estar en dos o más exclusiones. Luego de estas exclusiones, el número de predios en la base para trabajar es 705.025. En el cuadro 1 se muestran el número de predios excluidos en cada caso.

Cuadro 1: Número de predios en cada una de las exclusiones.

Exclusión	N
Rurales	97.321
Sin área	254.640
Sin Barmanpre	81.918
Sin estrato	19.043

### 3.1. Controles de calidad sobre la base

Debido a que la base de datos tiene información de diferentes fuentes y estas contienen información reportada directamente por los usuarios de plataformas digitales, puede darse el caso en que se tengan valores mal digitados o ciertas inconsistencias. Para evitar estos casos se propone realizar controles sobre esta información, con el objetivo de que la información utilizada por el modelo lineal sea lo más depurada posible.

La información a utilizar dentro de los diferentes modelos consiste en las áreas de construcción, precios reportados y los valores integrales. Considerando que se cuenta con datos, por un lado de ventas y arriendo, mientras que por otra parte predios en propiedad horizontal (PH) y no propiedad horizontal (NPH), además de diferentes estratos, se van a conformar diferentes agrupaciones de manera inicial. Las agrupaciones dependen del año de recolección del dato, clase de predio (PH y NPH), estrato y tipo de oferta (arriendo y venta). Es decir que, por ejemplo, las ofertas de arriendo en PH no se van a comparar con ofertas en arriendo en NPH, ni entre vigencias o entre estratos, ya que de antemano se espera que tengan un comportamiento diferente. Una vez se conforman estas agrupaciones, lo ideal es realizar comparaciones entre

### 3.1 *Controles de calidad sobre la base*

---

ofertas que sean próximas desde una perspectiva geográfica. Estas comparaciones se realizan con el objetivo de determinar si la oferta es similar en sus características a ofertas, debido a que en otro caso la información recolectada en ese registro presentaría anomalías y debería reportarse como atípica, al mismo tiempo que ser excluida de los análisis a desarrollar.

Una manera de determinar registros próximos geográficamente, es seleccionando todos los que se encuentren en el mismo sector catastral, pero un primer inconveniente surge cuando en un sector solamente se encuentre una oferta o un número insuficiente de ofertas para realizar la comparación. Por este motivo, se observa la necesidad de realizar aglomeraciones más grandes de tal forma que se tengan grupos con un número suficiente de información para realizar la validación. Es importante mencionar que si el número de registros supera el umbral determinado, no se requiere una agrupación más grande y se va a aplicar el protocolo de comparación únicamente con las ofertas que se encuentran en ese barrio.

Con el objetivo de tener agrupaciones de ofertas lo suficientemente grandes para tener información y realizar la comparación, pero de tal forma que no se agregue demasiado, considerando la singularidad de cada sector, en este documento se propone agrupar los sectores de acuerdo a su codificación. Se esperaría que los sectores cuya codificación es igual a diferencia del último dígito son más cercanos que dos sectores que tienen diferencias en sus dos últimos dígitos. Por ejemplo, se esperaría que los sectores “001101” y “001102” sean próximos geográficamente hablando. De esta manera, el algoritmo propuesto busca estas coincidencias a partir de la distancia de Levenshtein [Sariyar and Borg, 2020], la cual es un valor entre 0 y 1. Esta distancia toma un valor de 0 cuando las dos codificaciones no tienen dígitos en común en las mismas posiciones y toma el valor de 1 cuando son exactamente iguales. El algoritmo realiza la búsqueda iterativa de sectores similares, iniciando por los últimos dígitos, de forma que si no encuentra sectores con el mismo código y último dígito diferente, procede a buscar coincidencias exceptuando los dos últimos dígitos y así sucesivamente hasta superar el umbral definido. Una vez el algoritmo encuentra agrupaciones tal que todas superan dicho valor predefinido se detiene y no continúa con el proceso de agregación. El número de predios determinado, como umbral es 10, lo cual quiere decir que si un sector catastral cuenta con menos de 10 registros a nivel de tipo de oferta, estrato, clase de predio y año, se buscarán opciones de colapso, de tal forma que se encuentren ofertas próximas para poder realizar la comparación.

Una vez se conforman las diferentes agrupaciones se procede a utilizar una carta de control multivariada [Montgomery, 2020], lo que se debe a que se están controlando diferentes características, es decir un proceso multivariado, al involucrar el área construida, precio y el valor integral de las diferentes ofertas. Al considerar una carta de control multivariada se hace uso de la relación implícita dentro de las variables, lo cual no sucede con métodos univariados. Los métodos de control de calidad usualmente consideran dos etapas, lo que no se tiene en cuenta dentro de este ejercicio ya que no se tendrán en cuenta ofertas adicionales en este ejercicio y a por otra parte, el objetivo del estudio no se enfoca en el control de calidad, si no que éste es un paso adicional para lograr el objetivo, el cual radica en la estimación de la tasa de capitalización. La figura 1 muestra un ejemplo de una carta de control multivariada, donde las ofertas se ilustran a partir de puntos, las bandas determinan límites de tal forma que una oferta que está fuera de ellos será marcada como atípica y se muestran en el gráfico

### 3.1 Controles de calidad sobre la base

en color rojo.

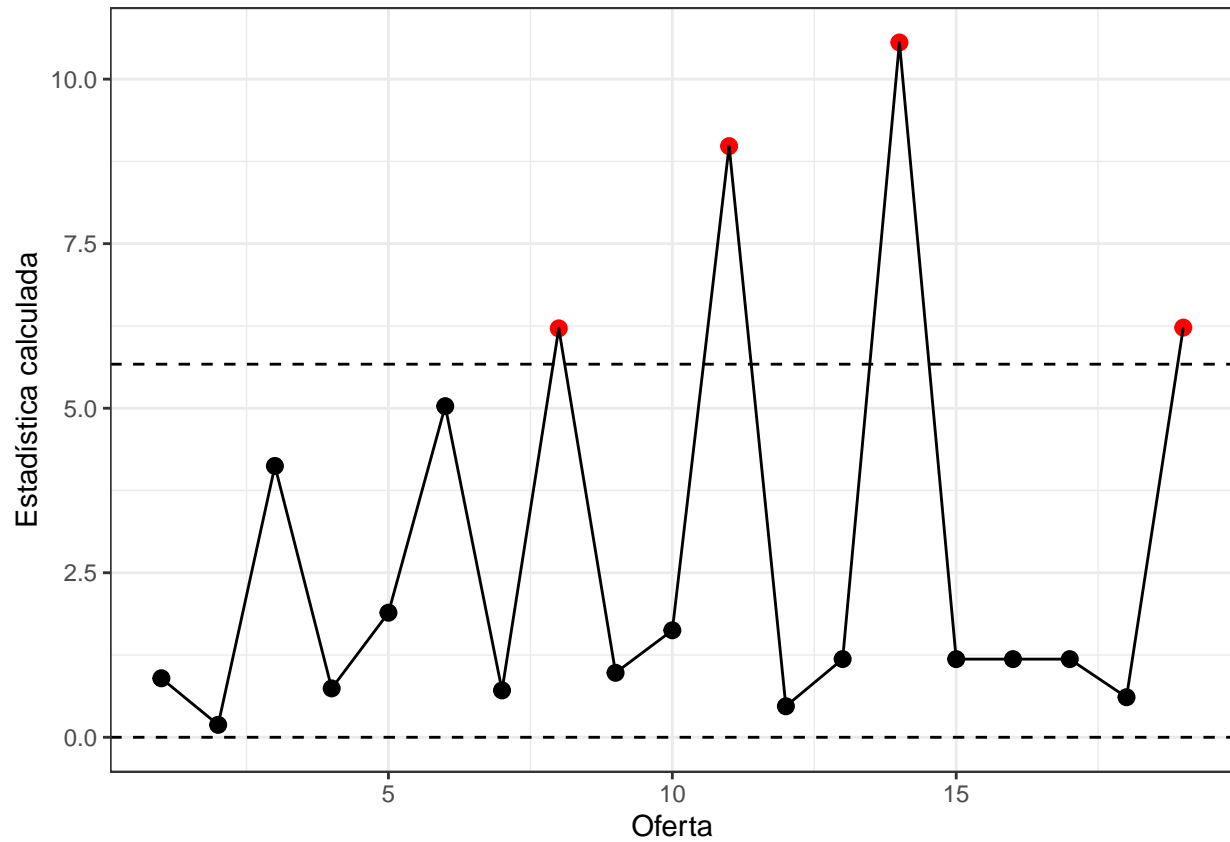


Figura 1: Ilustración de la carta de control multivariada.

Luego de este control de calidad, la base de datos tiene un total de 580.549 registros, con lo cual se procede a realizar los diferentes cálculos de promedios por sector, de tal manera que en cada uno de ellos se tendrá un promedio de venta y de arriendo, por estrato, vigencia y clase de predio (apartamento o casa). Al realizar este cálculo se tuvo un inconveniente, ya que un sector puede tener ofertas de arriendo, más no de ventas o viceversa. Para remediar esto, en el caso en que se tenga información de arriendo y no de ventas, se toma el promedio de valor integral de la base catastral, de acuerdo a la misma desagregación. Mientras que en el recíproco, cuando se tiene el valor de venta, pero no en el de arriendo se utiliza el promedio de la localidad respectivo para imputar este valor y proceder con la construcción del modelo estadístico.

Considerando que se elaboró una serie de procedimientos, controles de calidad, depuraciones e imputaciones se realizó un análisis para excluir registros que tuvieran tasas de renta atípicas, que puedan sesgar los resultados del estudio. Para esta situación, se excluyeron aquellos casos con una tasa de capitalización superior al 1 % o inferiores al 0.33 %. Luego de este proceso, la tabla 2 muestra el número de sectores resultantes, la cantidad de registros de ofertas y de ventas utilizadas en cada caso, respectivamente, donde se observa que la cantidad de ofertas

### 3.1 Controles de calidad sobre la base

de venta es mayor a la de arriendo y se encuentran cantidades mayores en estratos del 2 al 4.

Cuadro 2: Cantidad de predios y sectores según clase de predio, estrato y año, luego de la depuración de la base.

CLASE_PREDIO	AÑO	ESTRATO	N de sectores	N de ofertas de arriendo	N de ofertas de venta
N	2017	1	62	28	565
N	2017	2	188	227	3135
N	2017	3	224	748	5742
N	2017	4	70	580	1791
N	2017	5	30	227	737
N	2017	6	12	48	144
N	2018	1	36	29	310
N	2018	2	185	224	2960
N	2018	3	218	897	5520
N	2018	4	70	716	2334
N	2018	5	32	344	1097
N	2018	6	20	98	288
N	2019	1	51	29	540
N	2019	2	221	290	4453
N	2019	3	253	1235	8053
N	2019	4	77	1249	3019
N	2019	5	31	505	1386
N	2019	6	19	234	500
N	2020	1	53	27	559
N	2020	2	213	354	4052
N	2020	3	214	1117	7252
N	2020	4	77	1284	3030
N	2020	5	29	565	1727
N	2020	6	17	261	635
P	2017	1	34	107	129
P	2017	2	264	3393	5772
P	2017	3	355	12838	17683
P	2017	4	134	12567	17850
P	2017	5	57	5626	9903
P	2017	6	36	7591	11267
P	2018	1	36	99	137
P	2018	2	289	3486	5907
P	2018	3	363	14584	18694
P	2018	4	141	14470	20659

Cuadro 2: Cantidad de predios y sectores según clase de predio, estrato y año, luego de la depuración de la base.  
(continued)

CLASE_PREDIO	AÑO	ESTRATO	N de sectores	N de ofertas de arriendo	N de ofertas de venta
P	2018	5	59	7125	11711
P	2018	6	38	9113	13263
P	2019	1	31	90	103
P	2019	2	272	3046	5395
P	2019	3	369	12813	15742
P	2019	4	152	14826	20790
P	2019	5	65	7443	12099
P	2019	6	39	9699	13653
P	2020	1	34	97	98
P	2020	2	282	2998	4647
P	2020	3	369	14561	15157
P	2020	4	153	19044	20629
P	2020	5	64	9031	13356
P	2020	6	39	12894	16992
Total	Total	Total	6077	208857	331465

## 4. Modelo aplicado

El objetivo de esta sección es ilustrar el modelo estadístico aplicado junto con una breve justificación de su uso y respectivos resultados. Para eliminar el efecto que pueda tener el área construida sobre los valores de venta y de arriendo totales se va a tomar el valor integral, que se calcula dividiendo los valores totales sobre el área construida, es decir, lo que se muestra en la ecuación (2). En este modelo la variable respuesta es el valor integral de venta, mientras que una de las variables independientes es el valor integral de arriendo.

$$Valor\_integral = \frac{Valor\_total}{Area\_construida} \quad (2)$$

Considerando que no se cuenta con información de ventas y arriendo de manera simultánea en un solo predio, se toma la propuesta dada en Deaton [1985], cuyas cohortes van a estar representadas por los diferentes sectores catastrales. De esta manera, el modelo estimado se presenta en la ecuación (3), donde  $Z_{ij}$  y  $X_{ij}$  representan los valores integrales de venta y arriendo promedio, respectivamente, en el  $i$ -ésimo sector catastral en el  $j$ -ésimo tiempo,  $\alpha$  es el efecto global del valor integral de arriendo sobre el valor integral de venta. Por otro lado,  $\beta_k$  es el efecto del  $k$ -ésimo estrato sobre el efecto general de la relación entre valores



integrales, denotado como  $\alpha$ ,  $\eta_l$  es el efecto de la  $l$ -ésima localidad sobre el efecto general de la relación entre integrales y  $\delta_m$  es el efecto del  $m$ -ésimo sector catastral sobre el efecto global, de manera análoga a los otros coeficientes. Por último  $\epsilon_{ij}$  representa a los errores de la regresión.

$$Z_{ij} = \alpha X_{ij} + \beta_k X_{ij} + \eta_l X_{ij} + \delta_m X_{ij} + \epsilon_{ij} \quad (3)$$

En el caso del modelo (3), los factores relacionados al estrato y localidad son efectos fijos, ya que en la base de ofertas utilizada se tiene información de todos los estratos y de todas las localidades, mientras que los efectos relacionados con los sectores catastrales son efectos aleatorios, debido a que no se tienen mediciones de todos los sectores en la ciudad de Bogotá, pero la inferencia a realizar se puede generalizar a la población de sectores. Como en el modelo (3) se incluyen efectos fijos y aleatorios, el modelo se denomina de efectos mixtos.

En este modelo se supone lo siguiente

$$\sum_k \beta_k = \sum_l \eta_l = 0,$$

lo cual se debe a temas relacionados con la estimabilidad de los parámetros. Por otro lado, en cuanto a los factores aleatorios, se supone que los coeficientes y errores son independientes e idénticamente distribuidos y a su vez  $\delta_m \sim N(0, \sigma_\delta^2)$  y  $\epsilon_m \sim N(0, \sigma_\epsilon^2)$ . Es necesario tener en cuenta que el modelo no incluye intercepto, considerando los objetivos del estudio. El valor integral de arriendo, denotado por  $X_{ij}$  se incluye en todas las covariables, ya que el objetivo es determinar la relación entre los valores totales y de arriendo. Para esclarecer el proceso de estimación de la tasa de rentabilidad, el lector puede notar que la ecuación (3) se puede reexpresar como sigue

$$Z_{ij} = [\alpha + \beta_k + \eta_l + \delta_m] X_{ij} + \epsilon_{ij},$$

donde se observa que la tasa de renta, se va a calcular a partir de la suma de los respectivos coeficientes, según sea el caso.

Una vez se realizó la estimación de los coeficientes del modelo, se realizó un análisis para determinar puntos de apalancamiento o de influencia, con el fin de encontrar aquellos casos donde un solo registro pueda influir y cambiar los coeficientes del modelo de manera importante. Para ello se realizó el cálculo de la distancia de *Cook* y de los *DFBETAS*, que son medidas de diagnóstico del modelo. Detalles adicionales relacionados con modelos lineales de efectos mixtos o estas medidas de diagnóstico se pueden consultar en el anexo de este documento, en Montgomery et al. [2012], Hinkelmann [2011] o Melo et al. [2007].

A partir de este modelo se estima un coeficiente, mayor a 1, que nos permite conocer la razón entre el valor integral de venta y el valor integral o dicho de otra manera un valor que al multiplicarlo por el valor de arriendo, nos permita estimar el valor total de venta, pero para temas interpretativos con otras áreas de la Unidad e inclusive con la literatura relacionada,

## 4.1 Apartamentos

en los resultados presentados se va a mostrar el inverso de ese coeficiente, que va a ser un valor entre 0 y 1.

### 4.1. Apartamentos

A continuación se presentan los resultados de las diferentes tasas de rentabilidad para el caso de ofertas en propiedad horizontal. La figura 2 muestra diferentes diagramas de dispersión correspondientes a valores promedio de arriendo y de venta para los estratos en cada recuadro y diferentes años, representados por colores, para el caso de apartamentos. Allí se observa una relación lineal, lo cual justifica el uso del modelo (3) para este grupo de ofertas, al igual que una pendiente más empinada a medida que el estrato aumenta, lo que se refleja en una tasa de renta más baja para estratos altos y un comportamiento similar año a año.

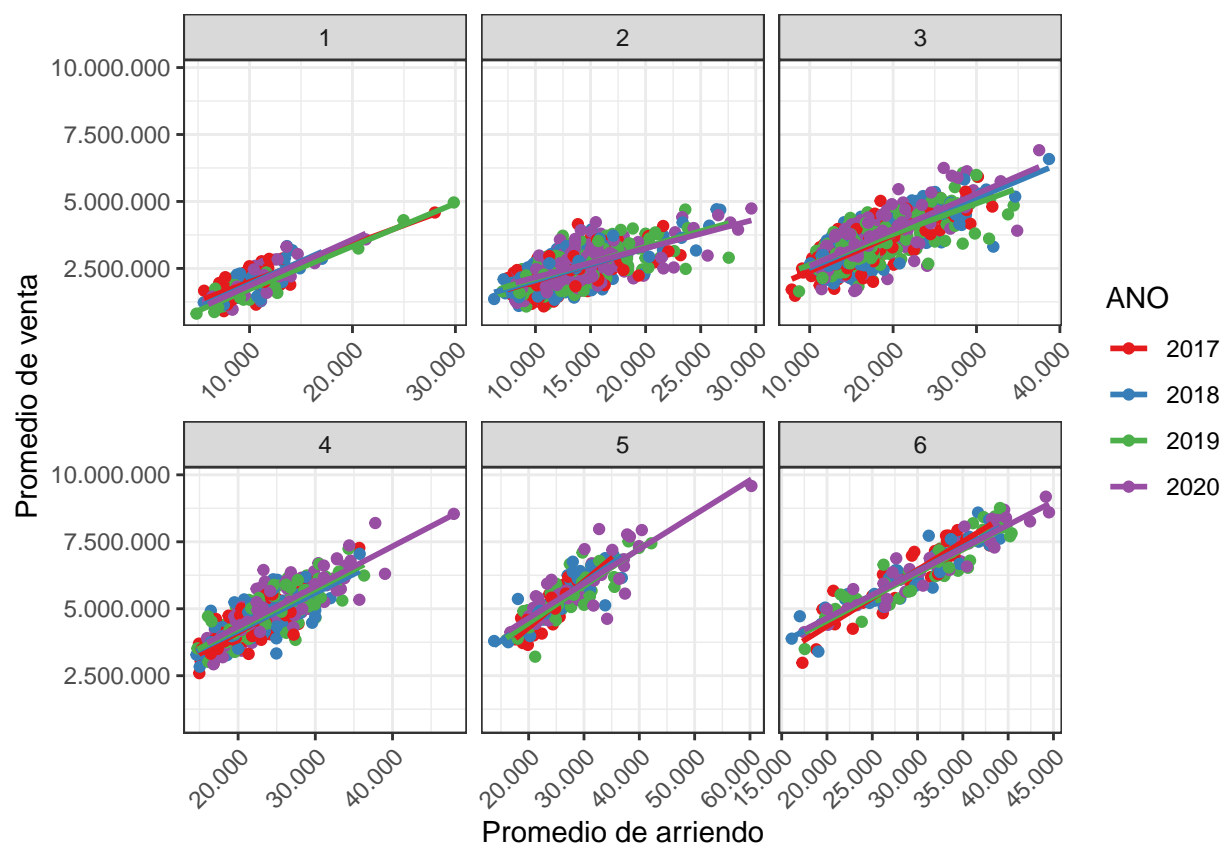


Figura 2: Diagrama de dispersión de los valores integrales de venta y de arriendo con sus respectivas rectas de regresión ajustadas por estrato y vigencia para apartamentos.

#### 4.1.1. Resultados

La tabla 3 y la tabla 4 muestran los resultados correspondientes a las tasas de renta a nivel de estrato y año, y a nivel de localidad y año. En el primer cuadro se observa que a medida que el estrato aumenta, la tasa de rentabilidad es menor. Las tasas de renta de los estratos 1

#### 4.1 Apartamentos

y 2 oscilan al rededor de 0.55 %. En el estrato 3 se tienen valores cercanos al 0.51 %, mientras que en el 4 y 5 se tienen valores cercanos al 0.49 % y por último en el 6 se tiene una tasa de renta cercana al 0.46 % a través de los diferentes años considerados. En el cuadro 4 se observa que las localidades con mayores tasas de renta son Chapinero, La Candelaria y San Cristobal. Por otro lado, la localidad con menor tasa de rentabilidad es Puente Aranda. Las figuras 3 y 4 muestran los mismos resultados de una manera gráfica.

Cuadro 3: Tasas de capitalización para apartamentos según estratos y años.

AÑO	ESTRATO	TASA_RENTA
2017	1	0.00540
2018	1	0.00538
2019	1	0.00546
2020	1	0.00540
2017	2	0.00553
2018	2	0.00550
2019	2	0.00559
2020	2	0.00553
2017	3	0.00516
2018	3	0.00514
2019	3	0.00521
2020	3	0.00515
2017	4	0.00491
2018	4	0.00489
2019	4	0.00496
2020	4	0.00491
2017	5	0.00483
2018	5	0.00482
2019	5	0.00488
2020	5	0.00483
2017	6	0.00459
2018	6	0.00457
2019	6	0.00463
2020	6	0.00459

Cuadro 4: Tasas de capitalización para apartamentos según localidades y años.

AÑO	CODIGO_LOCALIDAD	NOMBRE_LOCALIDAD	TASA_RENTA
2017	1	USAQUEN	0.00540
2018	1	USAQUEN	0.00538

#### 4.1 Apartamentos

Cuadro 4: Tasas de capitalización para apartamentos según localidades y años. *(continued)*

AÑO	CODIGO_LOCALIDAD	NOMBRE_LOCALIDAD	TASA_RENTA
2019	1	USAQUEN	0.00546
2020	1	USAQUEN	0.00540
2017	2	CHAPINERO	0.00579
2018	2	CHAPINERO	0.00576
2019	2	CHAPINERO	0.00585
2020	2	CHAPINERO	0.00578
2017	3	SANTA FE	0.00587
2018	3	SANTA FE	0.00584
2019	3	SANTA FE	0.00594
2020	3	SANTA FE	0.00586
2017	4	SAN CRISTOBAL	0.00549
2018	4	SAN CRISTOBAL	0.00547
2019	4	SAN CRISTOBAL	0.00555
2020	4	SAN CRISTOBAL	0.00549
2017	5	USME	0.00566
2018	5	USME	0.00563
2019	5	USME	0.00572
2020	5	USME	0.00565
2017	6	TUNJUELITO	0.00529
2018	6	TUNJUELITO	0.00527
2019	6	TUNJUELITO	0.00535
2020	6	TUNJUELITO	0.00529
2017	7	BOSA	0.00511
2018	7	BOSA	0.00509
2019	7	BOSA	0.00517
2020	7	BOSA	0.00511
2017	8	KENNEDY	0.00506
2018	8	KENNEDY	0.00504
2019	8	KENNEDY	0.00511
2020	8	KENNEDY	0.00506
2017	9	FONTIBON	0.00501
2018	9	FONTIBON	0.00499
2019	9	FONTIBON	0.00506
2020	9	FONTIBON	0.00501
2017	10	ENGATIVA	0.00506
2018	10	ENGATIVA	0.00503
2019	10	ENGATIVA	0.00511

#### 4.1 Apartamentos

Cuadro 4: Tasas de capitalización para apartamentos según localidades y años. *(continued)*

AÑO	CODIGO_LOCALIDAD	NOMBRE_LOCALIDAD	TASA_RENTA
2020	10	ENGATIVA	0.00505
2017	11	SUBA	0.00529
2018	11	SUBA	0.00527
2019	11	SUBA	0.00535
2020	11	SUBA	0.00529
2017	12	BARRIOS UNIDOS	0.00542
2018	12	BARRIOS UNIDOS	0.00539
2019	12	BARRIOS UNIDOS	0.00548
2020	12	BARRIOS UNIDOS	0.00542
2017	13	TEUSAQUILLO	0.00555
2018	13	TEUSAQUILLO	0.00552
2019	13	TEUSAQUILLO	0.00561
2020	13	TEUSAQUILLO	0.00555
2017	14	LOS MARTIRES	0.00547
2018	14	LOS MARTIRES	0.00544
2019	14	LOS MARTIRES	0.00553
2020	14	LOS MARTIRES	0.00547
2017	15	ANTONIO NARINO	0.00500
2018	15	ANTONIO NARINO	0.00498
2019	15	ANTONIO NARINO	0.00505
2020	15	ANTONIO NARINO	0.00499
2017	16	PUENTE ARANDA	0.00457
2018	16	PUENTE ARANDA	0.00456
2019	16	PUENTE ARANDA	0.00462
2020	16	PUENTE ARANDA	0.00457
2017	17	LA CANDELARIA	0.00585
2018	17	LA CANDELARIA	0.00582
2019	17	LA CANDELARIA	0.00592
2020	17	LA CANDELARIA	0.00584
2017	18	RAFAEL URIBE URIBE	0.00506
2018	18	RAFAEL URIBE URIBE	0.00504
2019	18	RAFAEL URIBE URIBE	0.00511
2020	18	RAFAEL URIBE URIBE	0.00505
2017	19	CIUDAD BOLIVAR	0.00545
2018	19	CIUDAD BOLIVAR	0.00543
2019	19	CIUDAD BOLIVAR	0.00551
2020	19	CIUDAD BOLIVAR	0.00545

## 4.1 Apartamentos

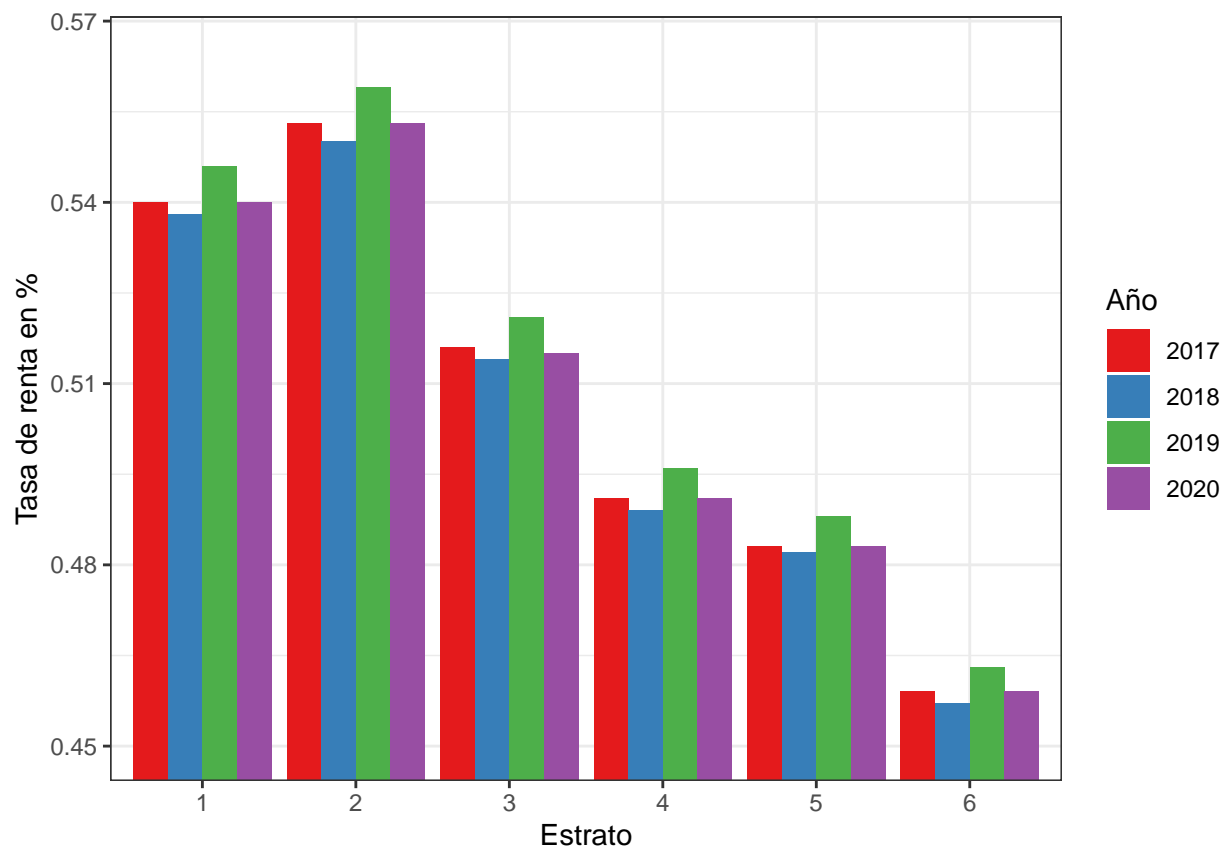


Figura 3: Tasas de capitalización para apartamentos según estratos y años.

#### 4.1 Apartamentos

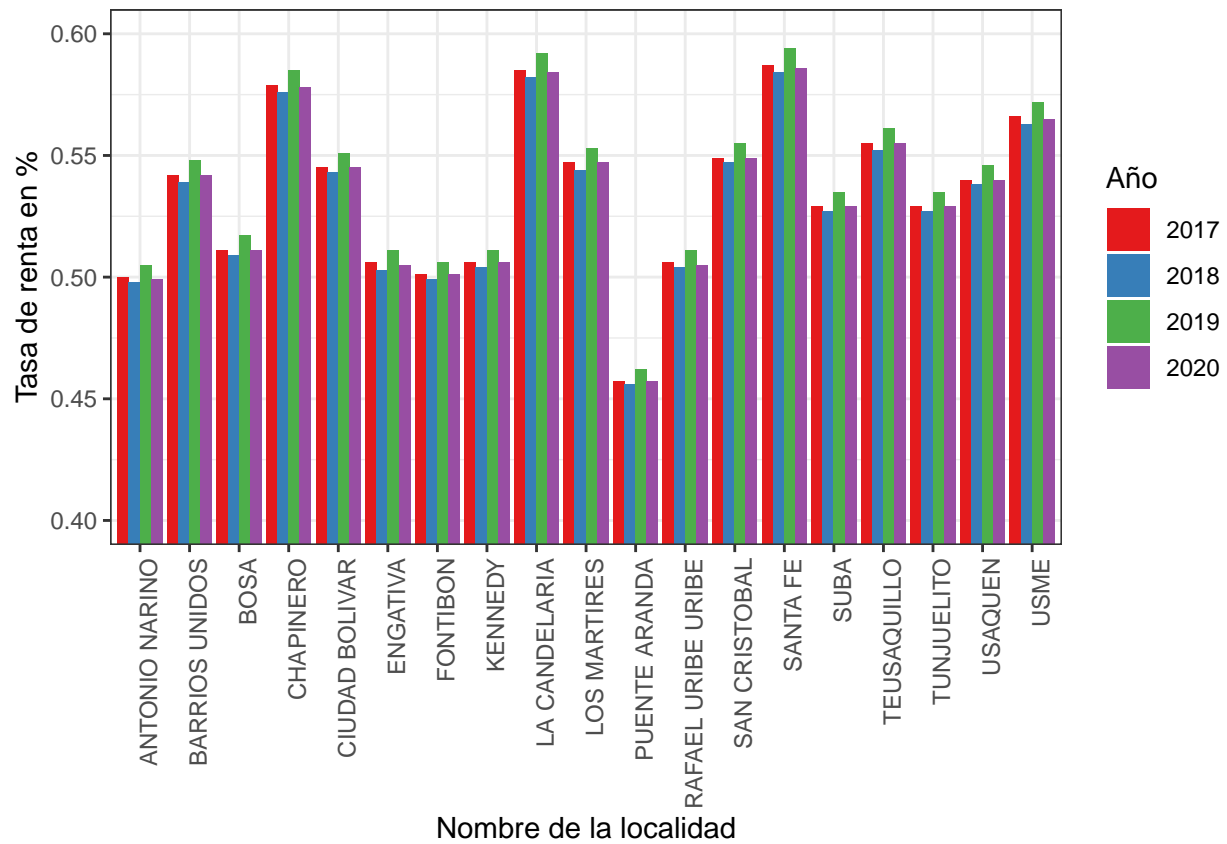
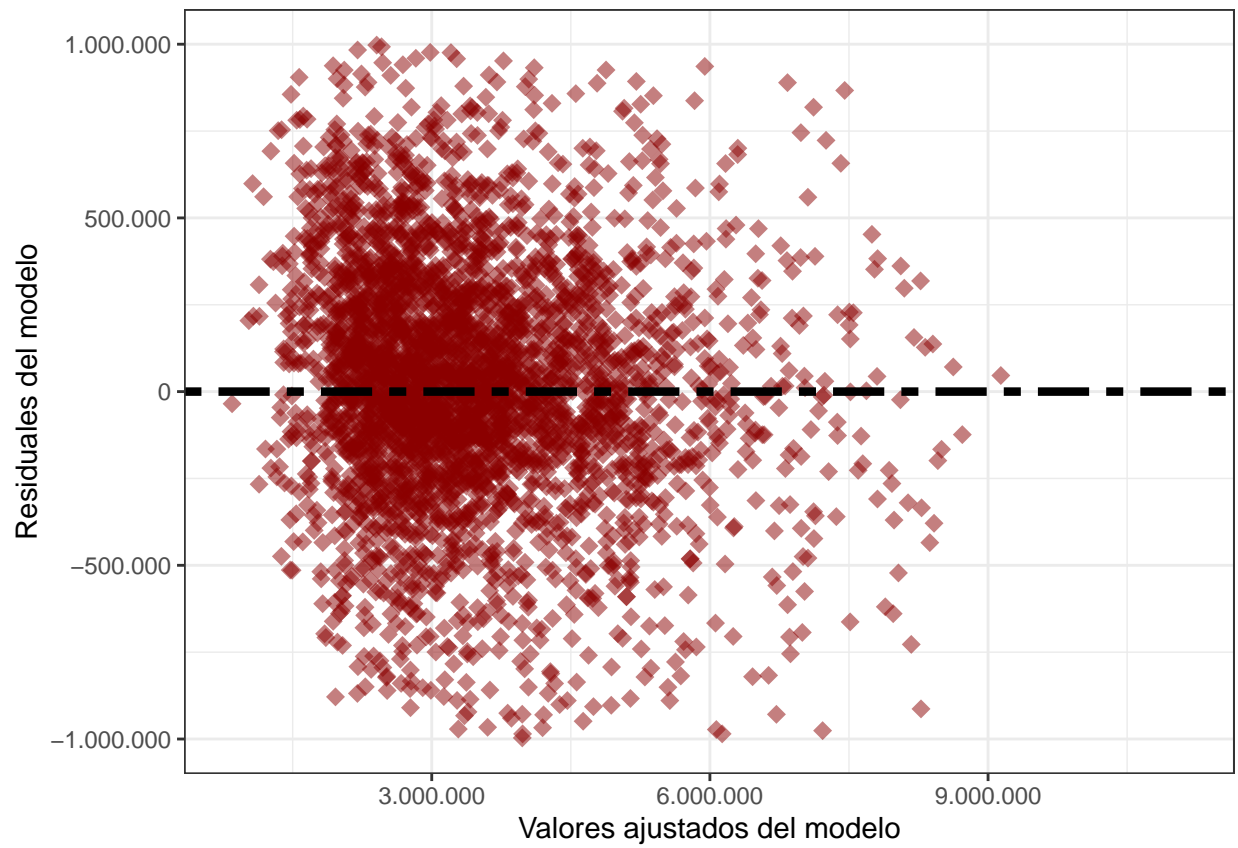


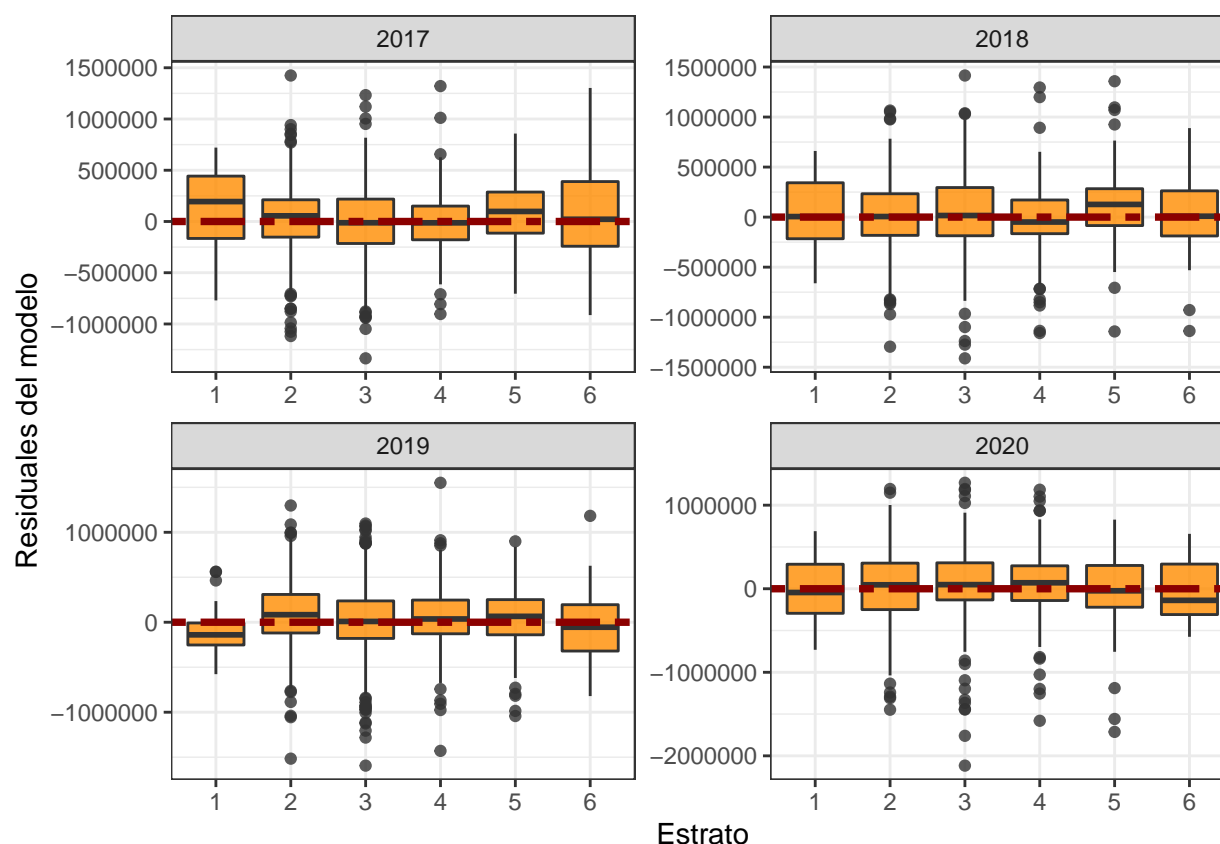
Figura 4: Tasas de capitalización para apartamentos según localidades y años.

#### 4.1 Apartamentos





## 4.2 Casas



### 4.1.2. Validación de supuestos

A manera de verificación de los supuestos del modelo para el grupo de ofertas de apartamentos, se realizó un análisis sobre los residuales del modelo. Las figuras ?? y ?? son gráficos de dispersión y de cajas y bigotes, que ilustran la relación entre los residuales con los valores ajustados, con los estratos y años, respectivamente. El objetivo de estas dos figuras es buscar patrones en los gráficos. Como se puede observar, no se encuentra una relación en los gráficos presentados.

Por otro lado, se validó el supuesto de normalidad incluido en los efectos aleatorios del modelo dado en la ecuación (3). Para ello, las figuras 5 y 6 muestran las densidades estimadas para los efectos aleatorios asociados a los sectores catastrales y a los residuales del modelo, respectivamente. En ambos casos se observa una forma acampanada, por lo cual se puede suponer que en ambos casos, los valores provienen de una distribución normal.

## 4.2. Casas

A continuación se presentan los resultados de las diferentes tasas de rentabilidad para el caso de ofertas en no propiedad horizontal. La figura 7 muestra diferentes diagramas de dispersión correspondientes a valores promedio de arriendo y de venta para los estratos en cada recuadro y diferentes años, representados por colores, para el caso de casas. Allí se observa una relación lineal, lo cual justifica el uso del modelo (3) para este grupo de ofertas, al igual que una

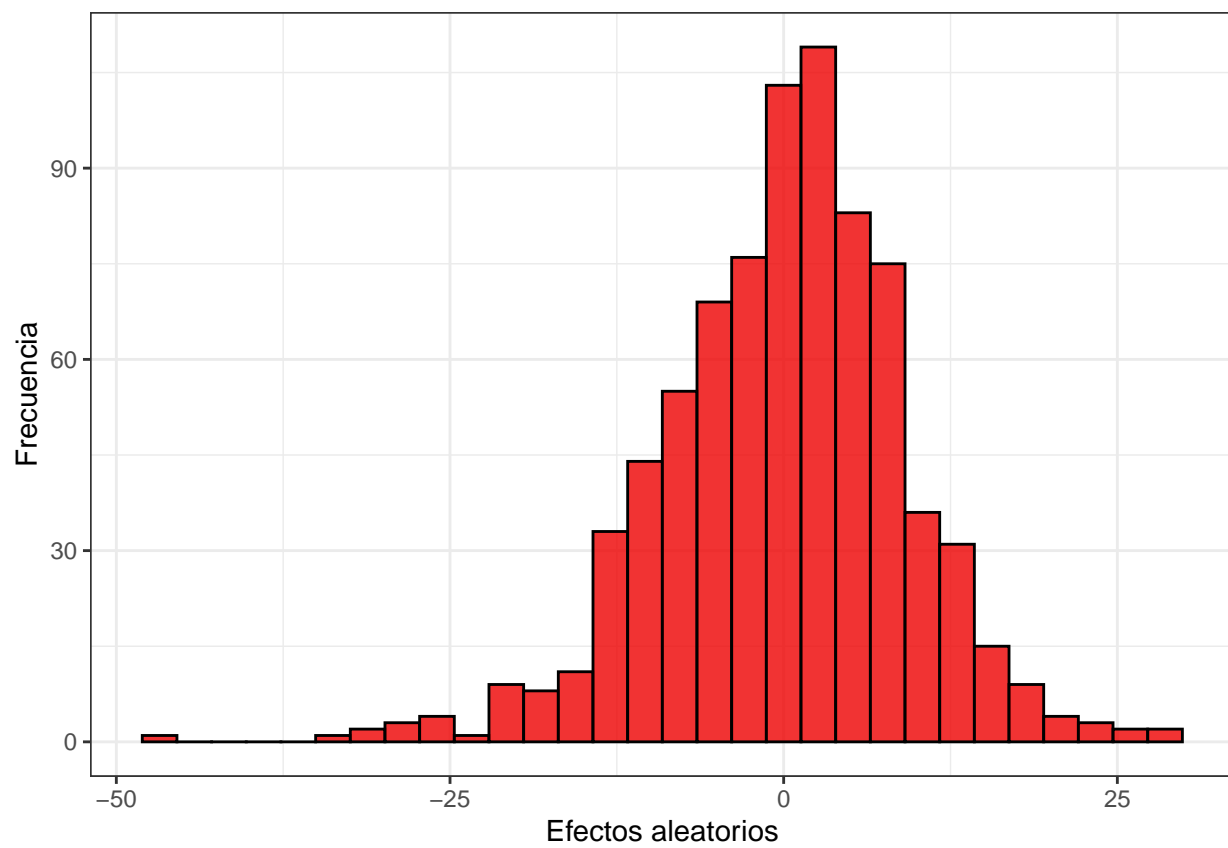


Figura 5: Densidad estimada de los efectos aleatorios asociados a los sectores en el modelo de apartamentos.

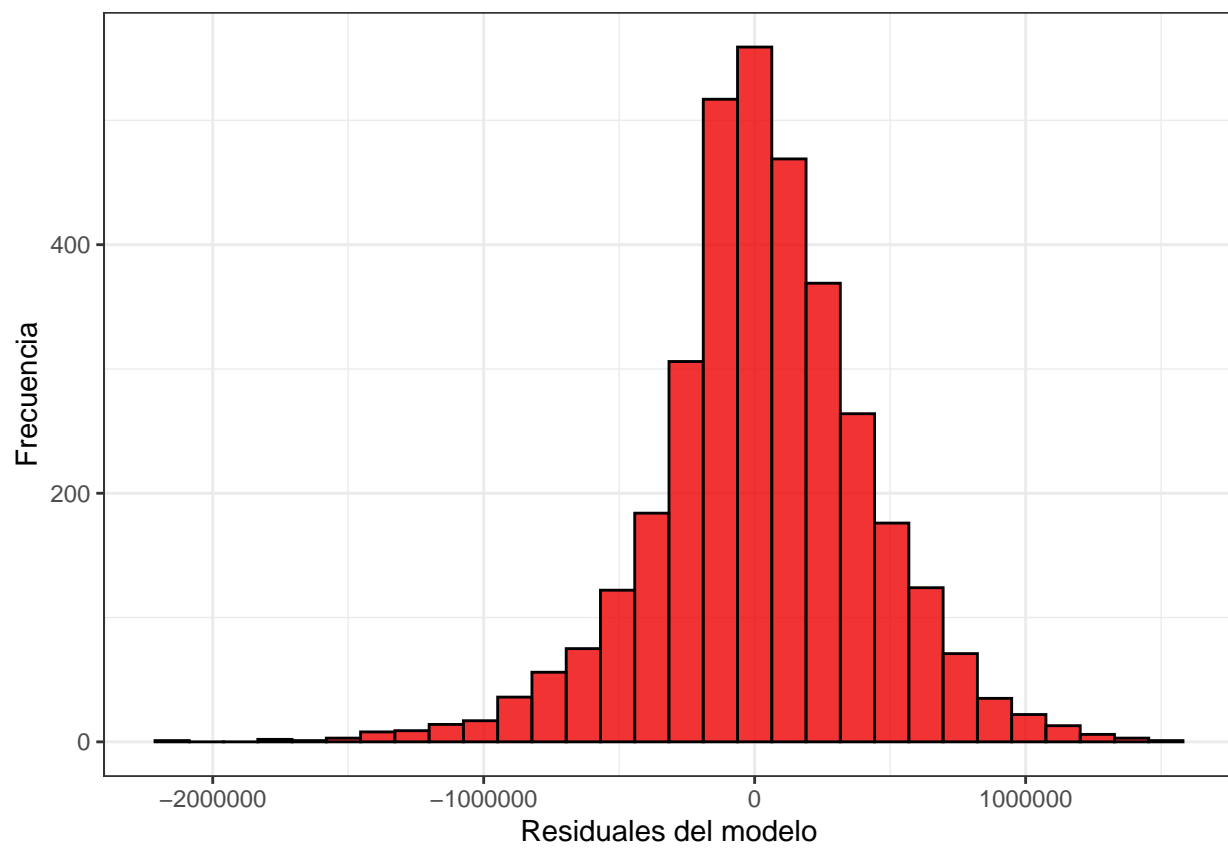


Figura 6: Densidad estimada de los residuales obtenidos en la regresión para el caso de apartamentos.

## 4.2 Casas

pendiente más empinada a medida que el estrato aumenta y un comportamiento similar año a través de los años.

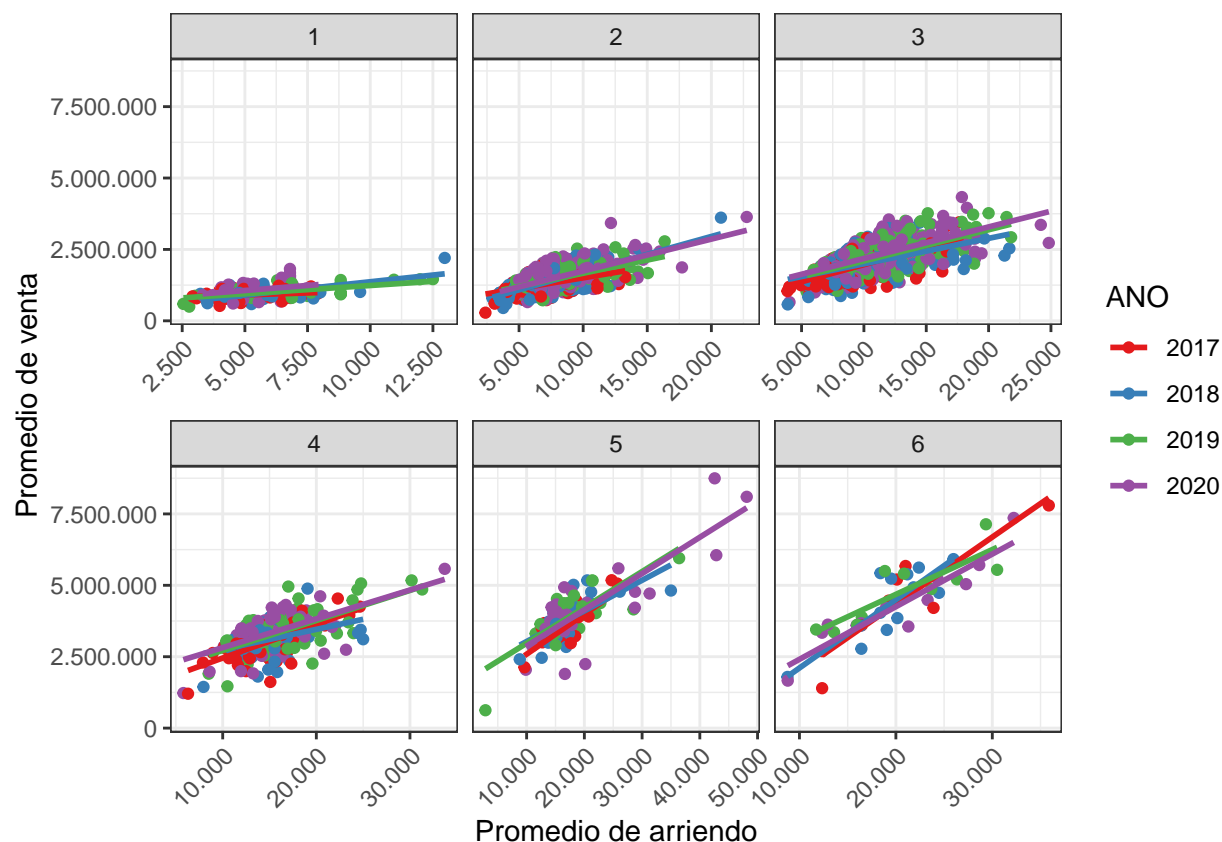


Figura 7: Diagrama de dispersión de los valores integrales de venta y de arriendo con sus respectivas rectas de regresión ajustadas por estrato y vigencia para apartamentos.

### 4.2.1. Resultados

El procedimiento en esta sección es análogo a lo presentado en el caso de propiedad horizontal. Las tablas 5 y 6 muestran los resultados para casas a nivel de estratos y localidades en las diferentes vigencias. Allí se observa, al igual que en el caso de la propiedad horizontal, que a medida que el estrato aumenta la tasa de rentabilidad es más baja. Las localidades con una mayor tasa de rentabilidad en casas son Chapinero, Los Mártires y Santa Fé, mientras que las de menor tasa son Bosa y Puente Aranda. Las figuras 8 y 9 muestran los mismos resultados de una manera gráfica. Para el caso de casas, a diferencia del de apartamentos, en el 2020 se observa una diferencia más notoria con respecto a los otros años, ya que en dicho año se tiene una tasa de renta más baja en comparación con los otros, mientras que en 2019 se tuvo la tasa más alta, seguida de 2017 y de 2018.

Cuadro 5: Tasas de capitalización para casas según estrato y años.

AÑO	ESTRATO	TASA_RENTA
2017	1	0.00590
2018	1	0.00580
2019	1	0.00597
2020	1	0.00569
2017	2	0.00532
2018	2	0.00524
2019	2	0.00538
2020	2	0.00515
2017	3	0.00491
2018	3	0.00484
2019	3	0.00496
2020	3	0.00476
2017	4	0.00448
2018	4	0.00442
2019	4	0.00452
2020	4	0.00435
2017	5	0.00453
2018	5	0.00447
2019	5	0.00457
2020	5	0.00440
2017	6	0.00443
2018	6	0.00438
2019	6	0.00447
2020	6	0.00431

Cuadro 6: Tasas de capitalización para casas según localidades y años.

AÑO	CODIGO_LOCALIDAD	NOMBRE_LOCALIDAD	TASA_RENTA
2017	1	USAQUEN	0.00590
2018	1	USAQUEN	0.00580
2019	1	USAQUEN	0.00597
2020	1	USAQUEN	0.00569
2017	2	CHAPINERO	0.00822
2018	2	CHAPINERO	0.00803
2019	2	CHAPINERO	0.00836
2020	2	CHAPINERO	0.00780

## 4.2 Casas

Cuadro 6: Tasas de capitalización para casas según localidades y años. *(continued)*

AÑO	CODIGO_LOCALIDAD	NOMBRE_LOCALIDAD	TASA_RENTA
2017	3	SANTA FE	0.00843
2018	3	SANTA FE	0.00823
2019	3	SANTA FE	0.00857
2020	3	SANTA FE	0.00799
2017	4	SAN CRISTOBAL	0.00671
2018	4	SAN CRISTOBAL	0.00659
2019	4	SAN CRISTOBAL	0.00680
2020	4	SAN CRISTOBAL	0.00643
2017	5	USME	0.00589
2018	5	USME	0.00579
2019	5	USME	0.00596
2020	5	USME	0.00567
2017	6	TUNJUELITO	0.00622
2018	6	TUNJUELITO	0.00611
2019	6	TUNJUELITO	0.00630
2020	6	TUNJUELITO	0.00598
2017	7	BOSA	0.00554
2018	7	BOSA	0.00545
2019	7	BOSA	0.00560
2020	7	BOSA	0.00535
2017	8	KENNEDY	0.00599
2018	8	KENNEDY	0.00589
2019	8	KENNEDY	0.00606
2020	8	KENNEDY	0.00577
2017	9	FONTIBON	0.00626
2018	9	FONTIBON	0.00614
2019	9	FONTIBON	0.00633
2020	9	FONTIBON	0.00601
2017	10	ENGATIVA	0.00588
2018	10	ENGATIVA	0.00579
2019	10	ENGATIVA	0.00595
2020	10	ENGATIVA	0.00567
2017	11	SUBA	0.00633
2018	11	SUBA	0.00621
2019	11	SUBA	0.00641
2020	11	SUBA	0.00608
2017	12	BARRIOS UNIDOS	0.00666

## 4.2 Casas

Cuadro 6: Tasas de capitalización para casas según localidades y años. *(continued)*

AÑO	CODIGO_LOCALIDAD	NOMBRE_LOCALIDAD	TASA_RENTA
2018	12	BARRIOS UNIDOS	0.00653
2019	12	BARRIOS UNIDOS	0.00675
2020	12	BARRIOS UNIDOS	0.00638
2017	13	TEUSAQUILLO	0.00698
2018	13	TEUSAQUILLO	0.00684
2019	13	TEUSAQUILLO	0.00708
2020	13	TEUSAQUILLO	0.00668
2017	14	LOS MARTIRES	0.00932
2018	14	LOS MARTIRES	0.00908
2019	14	LOS MARTIRES	0.00950
2020	14	LOS MARTIRES	0.00879
2017	15	ANTONIO NARINO	0.00643
2018	15	ANTONIO NARINO	0.00632
2019	15	ANTONIO NARINO	0.00652
2020	15	ANTONIO NARINO	0.00617
2017	16	PUENTE ARANDA	0.00551
2018	16	PUENTE ARANDA	0.00542
2019	16	PUENTE ARANDA	0.00557
2020	16	PUENTE ARANDA	0.00532
2017	17	LA CANDELARIA	0.00777
2018	17	LA CANDELARIA	0.00760
2019	17	LA CANDELARIA	0.00790
2020	17	LA CANDELARIA	0.00740
2017	18	RAFAEL URIBE URIBE	0.00619
2018	18	RAFAEL URIBE URIBE	0.00608
2019	18	RAFAEL URIBE URIBE	0.00627
2020	18	RAFAEL URIBE URIBE	0.00595
2017	19	CIUDAD BOLIVAR	0.00608
2018	19	CIUDAD BOLIVAR	0.00598
2019	19	CIUDAD BOLIVAR	0.00615
2020	19	CIUDAD BOLIVAR	0.00585

### 4.2.2. Validación de supuestos

A manera de verificación de los supuestos del modelo para el grupo de ofertas de apartamentos, se realizó un análisis sobre los residuales del modelo. Las figuras ?? y ?? son gráficos de dispersión y de cajas y bigotes, que ilustran la relación entre los residuales con los valores

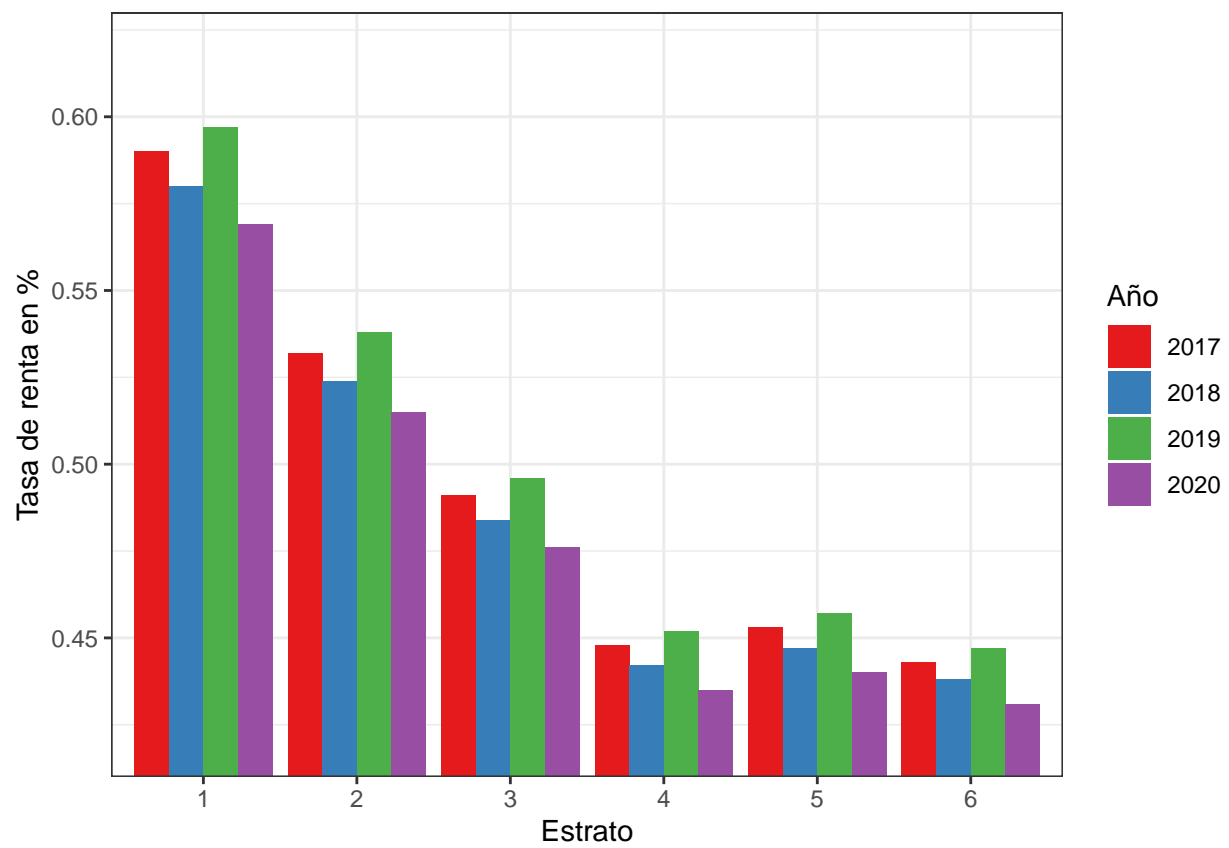


Figura 8: Tasas de capitalización para casas según estratos y años.



## 4.2 Casas

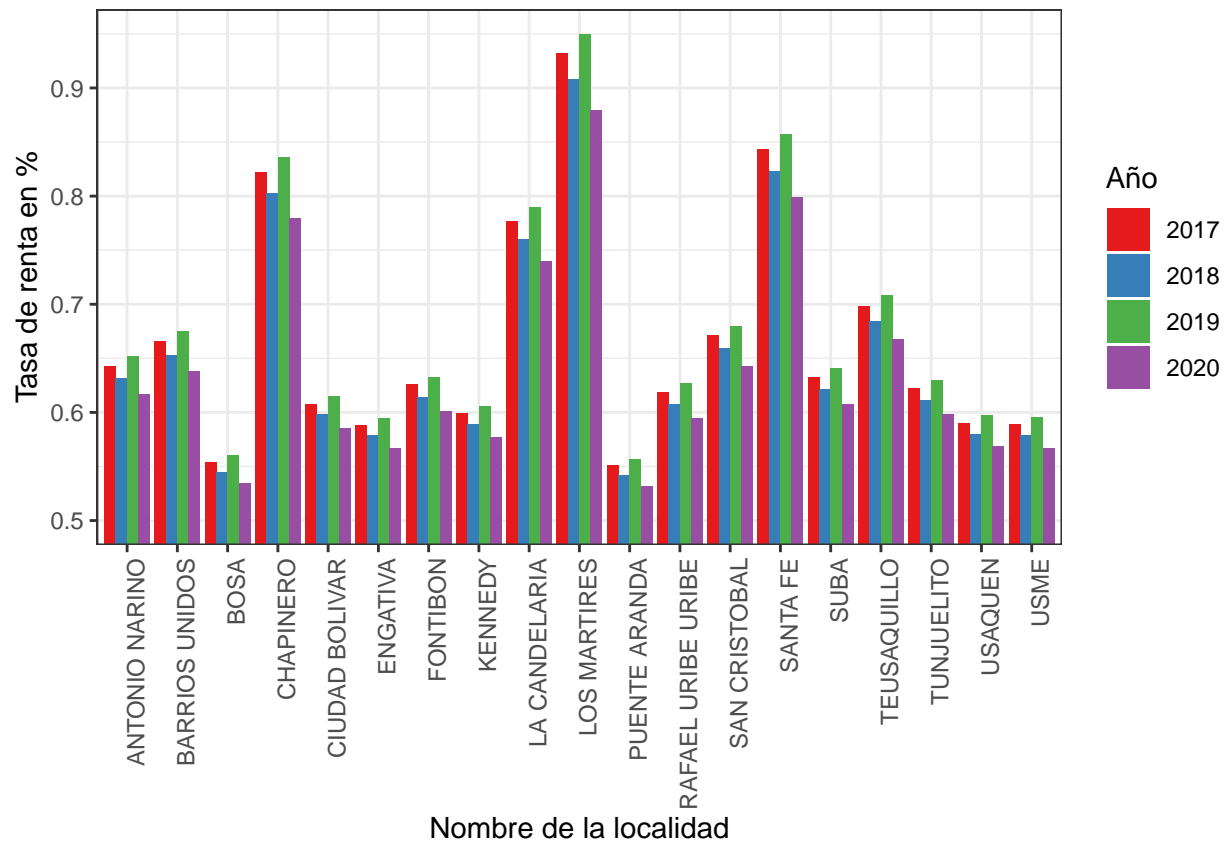
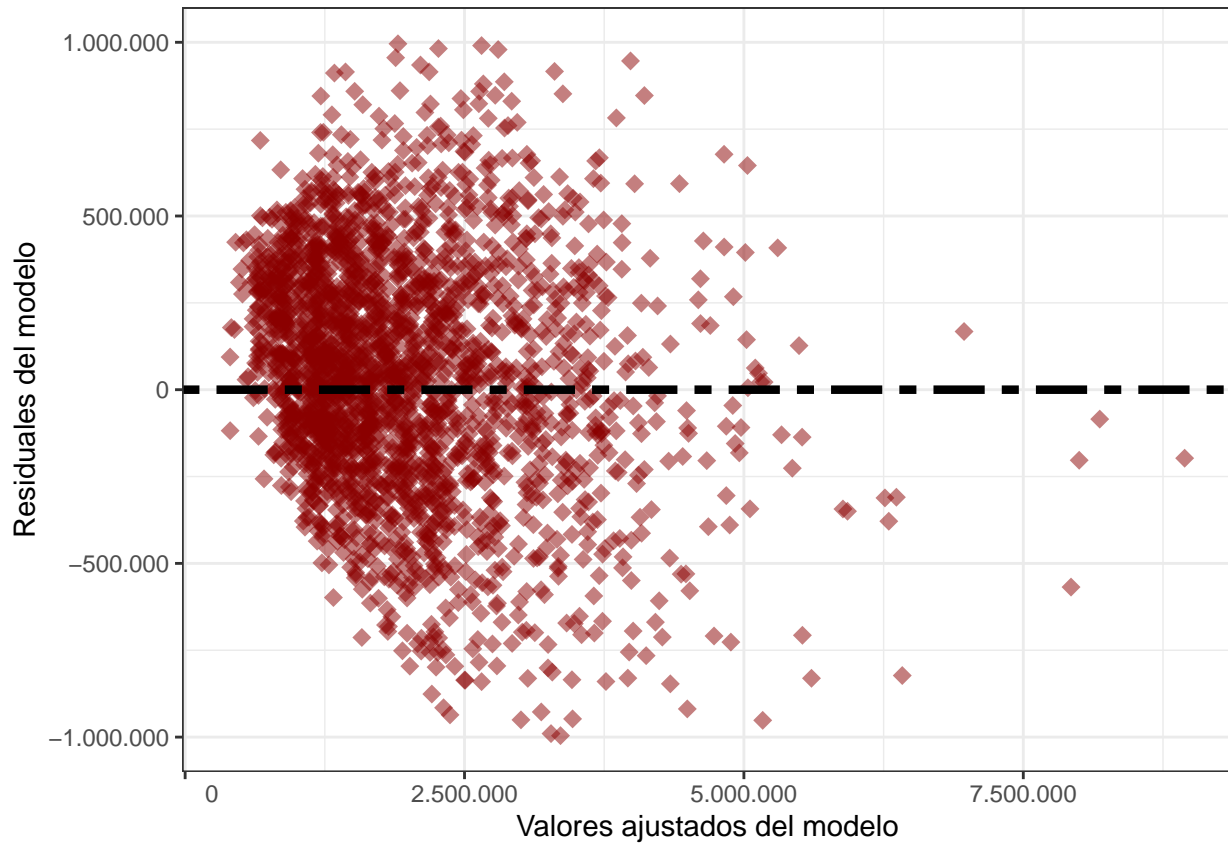
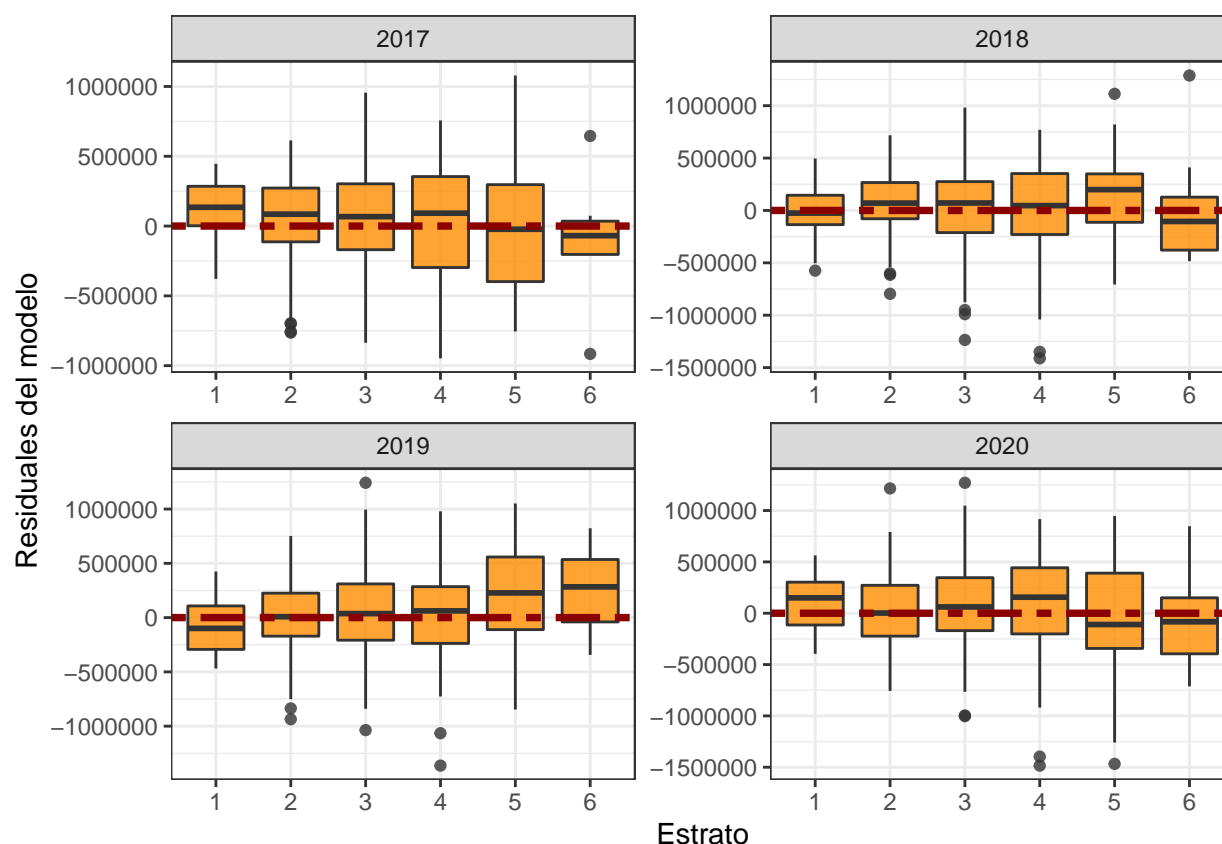


Figura 9: Tasas de capitalización para casas según localidades y años.

## 4.2 Casas

ajustrados, con los estratos y años, respectivamente. Por otro lado, se validó el supuesto de normalidad incluido en los efectos aleatorios del modelo dado en la ecuación (3). Para ello, las figuras 10 y 11 muestran las densidades estimadas para los efectos aleatorios asociados a los sectores catastrales y a los residuales del modelo, respectivamente. Las conclusiones de estos gráficos son las mismas a las presentadas para la validación de supuestos en apartamentos.





Es importante mencionar que las tasas de rentabilidad obtenidas para las ofertas en propiedad horizontal son mayores, por lo general a las que corresponden a no propiedad horizontal. Por último, los resultados obtenidos a nivel como sector se calcularon y hacen parte de los resultados de este documento, pero teniendo en cuenta que son tablas con un número de filas considerable, se omiten en este escrito.

## 5. Conclusiones

En este trabajo se realizó la estimación de la tasa de rentabilidad para predios residenciales a diferentes niveles de desagregación, a partir de métodos de regresión tradicionales. Dentro de los resultados de este documento, es importante resaltar las consideraciones realizadas en la depuración y limpieza de la información, con el fin de obtener una base de datos óptima, que permita realizar las estimaciones requeridas.

Los resultados obtenidos son a nivel de sector, año, clase de predio, localidad y estrato. En este documento se omiten los resultados a nivel de sector, debido a su tamaño. Las tasas de rentabilidad obtenidas para casas (NPH) son menores que las calculadas para apartamentos (PH). Se observa una relación inversamente proporcional entre el estrato y las tasas.

Para el caso NPH, las localidades con una mayor tasa de rentabilidad en casas son Chapinero, Los Mártires y Santa Fé, mientras que las de menor tasa son Bosa y Puente Aranda, mientras que en el caso PH las localidades con mayores tasas de renta son Chapinero, La Candelaria y

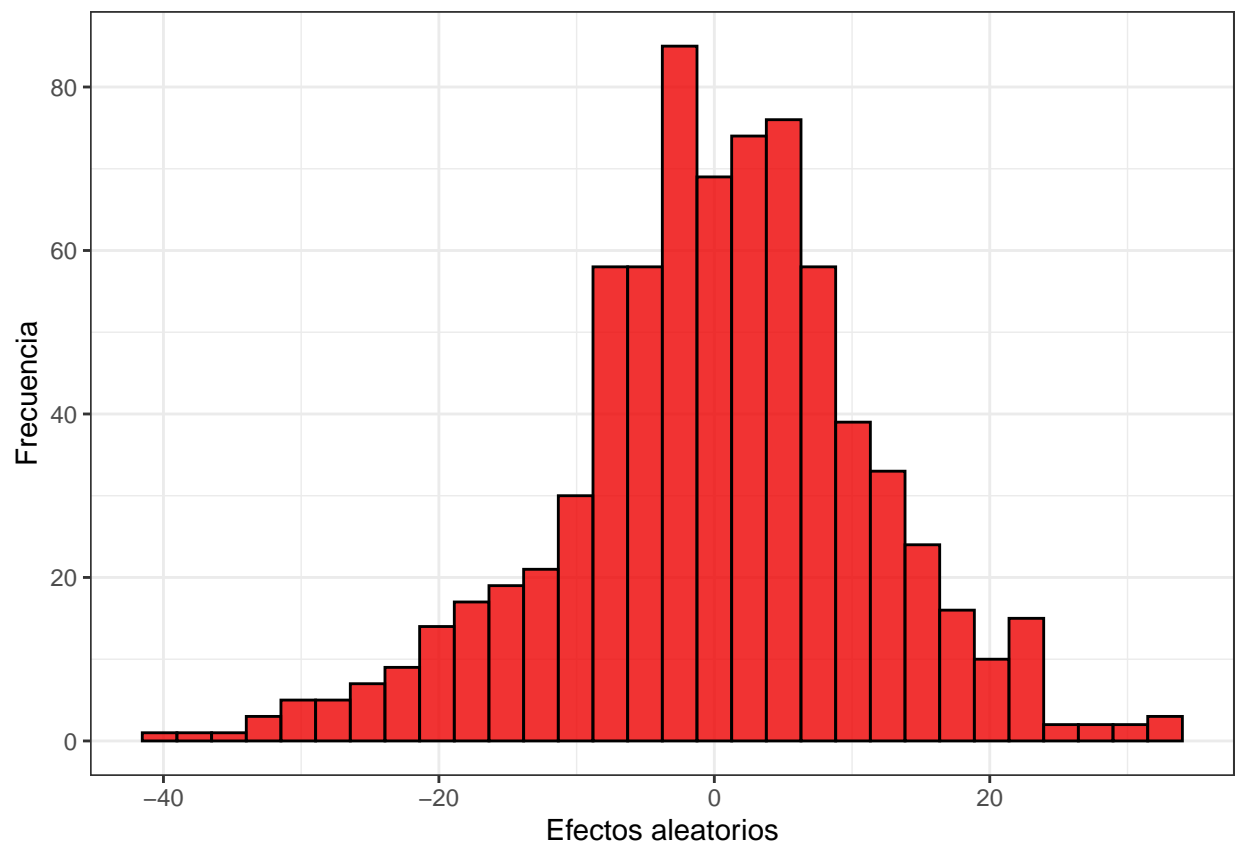


Figura 10: Densidad estimada de los efectos aleatorios asociados a los sectores en el modelo de casas.

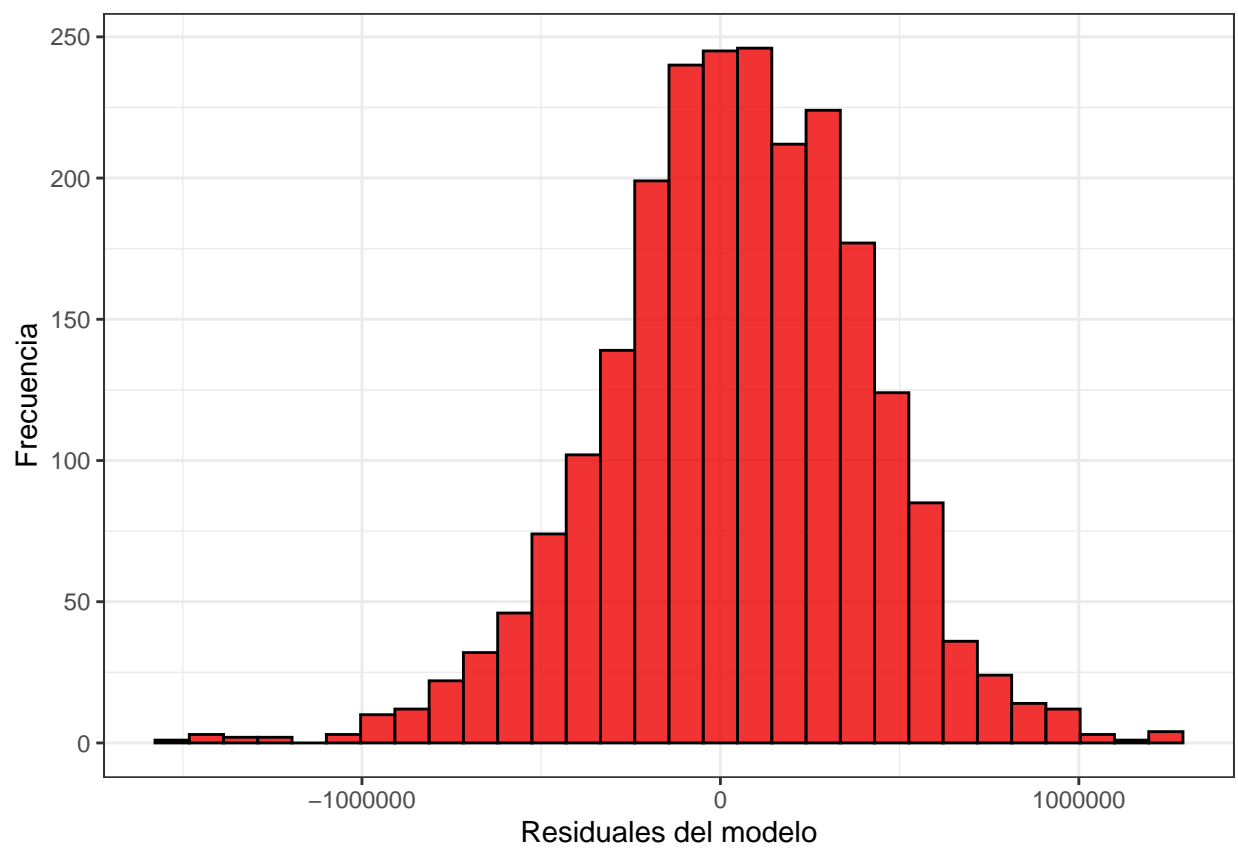


Figura 11: Densidad estimada de los residuales del modelo en el grupo de ofertas referente a casas.

San Cristobal. Por otro lado, la localidad con menor tasa de rentabilidad es Puente Aranda.

## 6. Trabajo futuro

- A partir de una depuración de la información, en futuros ejercicios se podría incluir la edad de los inmuebles como una variable independiente adicional en los modelos para obtener su tasa de rentabilidad asociada.
- Realizar estimaciones para usos diferentes al residencial.

## 7. Anexo 1

En esta sección se presenta una descripción, de manera corta, del modelo lineal de efectos mixtos, el cual fue utilizado en el documento para realizar las estimaciones requeridas dentro de los objetivos. Adicional a esta descripción se adicionan algunos criterios de validación que también fueron utilizados.

### 7.1. Modelo lineal mixto

Los modelos lineales son una forma de explicar la dispersión de una o más respuestas aleatorias en términos de una serie de variables independientes (también conocidas como exógenas). En West et al. [2014, p. 5] se encuentra un recuento histórico referente a los modelos lineales desde sus inicios en 1861 hasta la actualidad, desde una perspectiva teórica, mencionando al igual avances importantes en cuanto a la implementación de las diferentes metodologías a nivel de paquetes estadísticos. Éstos tienen en cuenta una serie de supuestos lo que hace posible su planteamiento, estimación, interpretación y su respectiva evaluación. Las variables independientes pueden ser clasificadas como efectos fijos o efectos aleatorios. Acorde con Melo et al. [2007, p. 6], cuando al finalizar el experimento las conclusiones se formulan sobre un número preestablecido de tratamientos el modelo se denomina de efectos fijos y en este caso la inferencia estadística se hace sobre los efectos medios de los tratamientos, por lo cual aquellas situaciones en que se desean realizar comparaciones o contrastes entre niveles de un factor en búsqueda de diferencias, éste se considera como fijo. Si los niveles de un atributo son una muestra aleatoria de una población de posibles selecciones, es decir, las conclusiones se formulan sobre un número mayor de tratamientos a los usados en el experimento, el modelo se dice de efectos aleatorios, y en este caso, la inferencia estadística se hace sobre las varianzas de los mismos. Los modelos que incluyen ambos se denominan de efectos mixtos (MLM).

En términos matriciales, según Pinheiro and Bates [2006, p. 58], el modelo lineal general con efectos mixtos viene dado de la forma

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \text{ con } i = 1, \dots, n \text{ y } n = r \times a \quad (4)$$

donde  $\mathbf{y}_i$  es el vector de respuestas,  $\mathbf{X}_i$  y  $\mathbf{Z}_i$  son las matrices de diseño de los efectos fijos y aleatorios asociadas al  $i$ -ésimo individuo, respectivamente. Los vectores  $\boldsymbol{\beta}$  y  $\mathbf{b}_i$  de dimensiones

## REFERENCIAS

---

$a$  y  $q$ , que contienen los coeficientes asociados a los efectos fijos y aleatorios considerados, respectivamente, mientras que  $\epsilon_i$  es el vector de errores de mediciones. Como supuestos distribucionales sobre este modelo se asume que  $\mathbf{b}_i \sim \mathbf{N}_q(\mathbf{0}_q, \Psi_{q \times q})$  y  $\epsilon_i \sim \mathbf{N}_t(\mathbf{0}_t, \Sigma_{t \times t})$ . A su vez los vectores  $\epsilon_i$  y  $\mathbf{b}_i$  se asumen independientes, es decir que  $Cov(\epsilon_i, \mathbf{b}_i) = 0$ . La matriz  $\Sigma$  asociada a los residuales dentro de las mediciones de la misma unidad puede considerar diferentes estructuras, es decir que se puede asociar a procesos autocorrelacionados, por ejemplo los que se encuentran en la literatura relacionada con series de tiempo, como los modelos *ARIMA* [Wei, 2006], o como los que se estudian en la estadística espacial, que se caracterizan mediante su función de variograma [Schabenberger and Gotway, 2017].

Las metodologías de estimación de los parámetros incluidos dentro del modelo son máxima verosimilitud o máxima verosimilitud restringida, a pesar de que en la literatura se encuentran diferentes propuestas. En este documento se utilizó la primera, ya que es la que se incluye por defecto dentro de la función *lme* del paquete *nlme* [Pinheiro et al., 2021] en R.

## Referencias

- Francisca Antman and David J McKenzie. Earnings mobility and measurement error: A pseudo-panel approach. *Economic Development and Cultural Change*, 56(1):125–161, 2007.
- Angus Deaton. Panel data from time series of cross-sections. *Journal of Econometrics*, 30(1-2):109–126, 1985.
- Klaus Hinkelmann. *Design and analysis of experiments, special designs and applications*, volume 3. John Wiley and Sons, Blacksburg, 2011.
- O Melo, L López, and S Melo. Diseño de experimentos: métodos y aplicaciones. *Editorial Universidad Nacional de Colombia*. Bogotá, 2007.
- Douglas C Montgomery. *Introduction to statistical quality control*. John Wiley & Sons, 2020.
- Douglas C Montgomery, Elizabeth A Peck, and G Geoffrey Vining. *Introduction to Linear Regression Analysis*, volume 821. John Wiley and Sons, New York, 2012.
- José Pinheiro and Douglas Bates. *Mixed-effects models in S and S-PLUS*. Springer Science & Business Media, New York, 2006.
- Jose Pinheiro, Douglas Bates, Saikat DebRoy, Deepayan Sarkar, and R Core Team. *nlme: Linear and Nonlinear Mixed Effects Models*, 2021. URL <https://CRAN.R-project.org/package=nlme>. R package version 3.1-152.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021. URL <https://www.R-project.org/>.
- Murat Sariyar and Andreas Borg. *RecordLinkage: Record Linkage Functions for Linking and Deduplicating Data Sets*, 2020. URL <https://CRAN.R-project.org/package=RecordLinkage>. R package version 0.4-12.1.
- Oliver Schabenberger and Carol A Gotway. *Statistical methods for spatial data analysis*. Chapman and Hall/CRC, Boca Raton, 2017.

## REFERENCIAS

---

- Luca Scrucca. qcc: an r package for quality control charting and statistical process control. *R News*, 4/1:11–17, 2004. URL <https://cran.r-project.org/doc/Rnews/>.
- Maresa Sprietsma. Computers as pedagogical tools in Brazil: a pseudo-panel analysis. *Education Economics*, 20(1):19–32, 2012.
- Ainhoa Oguiza Tovar, Inmaculada Gallastegui Zulaica, and Vicente Núñez-Antón. Analysis of pseudo-panel data with dependent samples. *Journal of Applied Statistics*, 39(9):1921–1937, 2012.
- Chi-Hong Tsai, Corinne Mulley, and Geoffrey Clifton. A review of pseudo panel data approach in estimating short-run and long-run public transport demand elasticities. *Transport Reviews*, 34(1):102–121, 2014.
- William WS Wei. Time series analysis. In *The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2*. 2006.
- Brady T West, Kathleen B Welch, and Andrzej T Galecki. *Linear mixed models: a practical guide using Statistical Software*. CRC Press, Boca Raton, 2014.
- Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Golemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani. Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686, 2019. doi: 10.21105/joss.01686.