

Metodología para la estimación de la tasa de capitalización

Grupo estadístico - Observatorio técnico catastral (UAECD)

Contents

1	Introducción	1
2	Objetivos	2
2.1	Objetivo general	2
2.2	Objetivos específicos	2
3	Marco teórico	3
4	Modelo lineal mixto	3
5	Revisión de la base de ofertas	3

Abstract

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 la secundaria corresponde 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 para estimar 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 β , 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 rudimentos correspondientes a las metodologías de modelos lineales [Melo et al., 2007] que permitan estimar el coeficiente β . Este ejercicio se debe realizar para diferentes desagregaciones de predios en la ciudad, tales como usos, estratos y/o desagregaciones geográficas (localidad, upz, sectores catastrales, entre otros). El nivel de desagregación depende de la cantidad de registros correspondiente a cada caso.

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

En la información disponible para la realización de este documento, se encuentran las ofertas recolectadas por la UAECD tanto de fuentes primarias como secundarias en los años, 2017, 2018, 2019 y 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.

Dentro de la literatura se encuentran diversos ejemplos de este tipo 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. 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 fijos correspondientes a cada una de las diferentes cohortes. El ejemplo que se trabaja allí es la creación de las agrupaciones mediante la fecha de nacimiento de las personas, de tal forma que cada uno pertenece únicamente a un grupo en el análisis realizado. 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).

A partir de la propuesta de Deaton [1985] se particiona la población en las cohortes mencionadas y considerando que dichos segmentos abarcan todo el conjunto, los efectos incluidos en los modelos lineales de regresión son efectos fijos, a diferencia del caso que se trabaja en este documento donde se utiliza el modelo propuesto allí, pero además se incluyen efectos aleatorios asociados a las diferentes cohortes, debido a que generalmente se trabaja únicamente con un subconjunto de UE pertenecientes a la población objetivo y no la totalidad de unidades.

Por ende, dado que no se tienen mediciones de arriendo y venta sobre los mismos predios, se propone el uso de un modelo lineal, agregando de acuerdo a “cohortes”, de tal manera que se pueda emular la propuesta de Deaton [1985]. En este caso, la propuesta es agrupar por agregaciones geográficas tales como sectores catastrales, considerando que desagregar a nivel de lote puede llegar a ser demasiado fino y pueda presentar problemas a la hora de hacer la estimación de parámetros en el modelo lineal, y por otro lado a nivel de agregaciones como localidad puede generar agrupaciones muy grandes, considerando los objetivos del estudio.

El presente documento contiene en la sección la introducción, seguido de la sección donde se presentan los objetivos del estudio. En la sección 3 se hace una revisión metodológica de las metodologías estadísticas a utilizar. Por otro lado, la sección

2 Objetivos

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

3 Marco teórico

4 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

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \text{ con } i = 1, \dots, n \text{ y } n = r \times a \quad (2)$$

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 a y q , que contienen los coeficientes asociados a los efectos fijos y aleatorios considerados en la i -ésima UE, respectivamente, mientras que $\boldsymbol{\epsilon}_i$ es el vector de errores de mediciones dentro de cada UE. Como supuestos distribucionales sobre este modelo se asume que $\mathbf{b}_i \sim N_q(\mathbf{0}_q, \boldsymbol{\Psi}_{q \times q})$ y $\boldsymbol{\epsilon}_i \sim N_t(\mathbf{0}_t, \boldsymbol{\Sigma}_{t \times t})$. A su vez los vectores $\boldsymbol{\epsilon}_i$ y \mathbf{b}_i se asumen independientes, es decir $Cov(\boldsymbol{\epsilon}_i, \mathbf{b}_i) = 0$. La matriz $\boldsymbol{\Sigma}$ asociada a los residuales dentro de las mediciones de la misma unidad puede considerar diferentes situaciones, 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]. En este documento únicamente aborda el contexto de datos longitudinales o datos panel, por lo cual únicamente se tendrán en cuenta aquellos procesos relacionados con series temporales en cuanto a la matriz de errores $\boldsymbol{\Sigma}$.

5 Revisión de la base de ofertas

A	ESTRATO	TASA_RENTA	TASA_RENTA_2
2017	1	185.67	0.0054
2018	1	186.18	0.0054
2019	1	183.64	0.0054
2020	1	186.06	0.0054
2017	2	181.87	0.0055
2018	2	182.39	0.0055
2019	2	179.85	0.0056
2020	2	182.27	0.0055
2017	3	194.93	0.0051
2018	3	195.44	0.0051
2019	3	192.91	0.0052

(continued)

A	ESTRATO	TASA_RENTA	TASA_RENTA_2
2020	3	195.33	0.0051
2017	4	204.05	0.0049
2018	4	204.56	0.0049
2019	4	202.03	0.0049
2020	4	204.45	0.0049
2017	5	208.43	0.0048
2018	5	208.95	0.0048
2019	5	206.41	0.0048
2020	5	208.83	0.0048
2017	6	217.22	0.0046
2018	6	217.74	0.0046
2019	6	215.20	0.0046
2020	6	217.62	0.0046

A	CODIGO_LOCALIDAD	TASA_RENTA	TASA_RENTA_2
2017	1	185.67	0.0054
2018	1	186.18	0.0054
2019	1	183.64	0.0054
2020	1	186.06	0.0054
2017	2	174.08	0.0057
2018	2	174.59	0.0057
2019	2	172.05	0.0058
2020	2	174.47	0.0057
2017	3	168.94	0.0059
2018	3	169.46	0.0059
2019	3	166.92	0.0060
2020	3	169.34	0.0059
2017	4	182.30	0.0055
2018	4	182.81	0.0055
2019	4	180.28	0.0055
2020	4	182.70	0.0055
2017	5	177.41	0.0056
2018	5	177.93	0.0056
2019	5	175.39	0.0057
2020	5	177.81	0.0056
2017	6	188.62	0.0053
2018	6	189.14	0.0053
2019	6	186.60	0.0054
2020	6	189.02	0.0053
2017	7	195.27	0.0051
2018	7	195.78	0.0051
2019	7	193.25	0.0052
2020	7	195.67	0.0051
2017	8	197.28	0.0051
2018	8	197.79	0.0051
2019	8	195.25	0.0051
2020	8	197.68	0.0051

(continued)

A	CODIGO_LOCALIDAD	TASA_RENTA	TASA_RENTA_2
2017	9	199.23	0.0050
2018	9	199.75	0.0050
2019	9	197.21	0.0051
2020	9	199.63	0.0050
2017	10	197.22	0.0051
2018	10	197.73	0.0051
2019	10	195.20	0.0051
2020	10	197.62	0.0051
2017	11	188.62	0.0053
2018	11	189.13	0.0053
2019	11	186.59	0.0054
2020	11	189.02	0.0053
2017	12	184.33	0.0054
2018	12	184.84	0.0054
2019	12	182.30	0.0055
2020	12	184.72	0.0054
2017	13	180.05	0.0056
2018	13	180.56	0.0055
2019	13	178.02	0.0056
2020	13	180.44	0.0055
2017	14	181.69	0.0055
2018	14	182.21	0.0055
2019	14	179.67	0.0056
2020	14	182.09	0.0055
2017	15	200.15	0.0050
2018	15	200.66	0.0050
2019	15	198.13	0.0050
2020	15	200.55	0.0050
2017	16	218.45	0.0046
2018	16	218.96	0.0046
2019	16	216.42	0.0046
2020	16	218.84	0.0046
2017	17	170.27	0.0059
2018	17	170.79	0.0059
2019	17	168.25	0.0059
2020	17	170.67	0.0059
2017	18	197.57	0.0051
2018	18	198.09	0.0050
2019	18	195.55	0.0051
2020	18	197.97	0.0051
2017	19	183.47	0.0055
2018	19	183.99	0.0054
2019	19	181.45	0.0055
2020	19	183.87	0.0054

A	ESTRATO	TASA_RENTA	TASA_RENTA_2
2017	1	168.24	0.0059
2018	1	174.02	0.0057
2019	1	173.79	0.0058
2020	1	180.94	0.0055
2017	2	179.43	0.0056
2018	2	185.20	0.0054
2019	2	184.98	0.0054
2020	2	192.13	0.0052
2017	3	196.41	0.0051
2018	3	202.18	0.0049
2019	3	201.95	0.0050
2020	3	209.10	0.0048
2017	4	215.75	0.0046
2018	4	221.53	0.0045
2019	4	221.30	0.0045
2020	4	228.45	0.0044
2017	5	213.74	0.0047
2018	5	219.52	0.0046
2019	5	219.29	0.0046
2020	5	226.44	0.0044
2017	6	221.73	0.0045
2018	6	227.50	0.0044
2019	6	227.28	0.0044
2020	6	234.42	0.0043

A	CODIGO_LOCALIDAD	TASA_RENTA	TASA_RENTA_2
2017	1	168.24	0.0059
2018	1	174.02	0.0057
2019	1	173.79	0.0058
2020	1	180.94	0.0055
2017	2	125.19	0.0080
2018	2	130.96	0.0076
2019	2	130.73	0.0076
2020	2	137.88	0.0073
2017	3	132.31	0.0076
2018	3	138.08	0.0072
2019	3	137.85	0.0073
2020	3	145.00	0.0069
2017	4	149.13	0.0067
2018	4	154.91	0.0065
2019	4	154.68	0.0065
2020	4	161.83	0.0062
2017	5	162.98	0.0061
2018	5	168.75	0.0059
2019	5	168.52	0.0059
2020	5	175.67	0.0057
2017	6	165.26	0.0061

(continued)

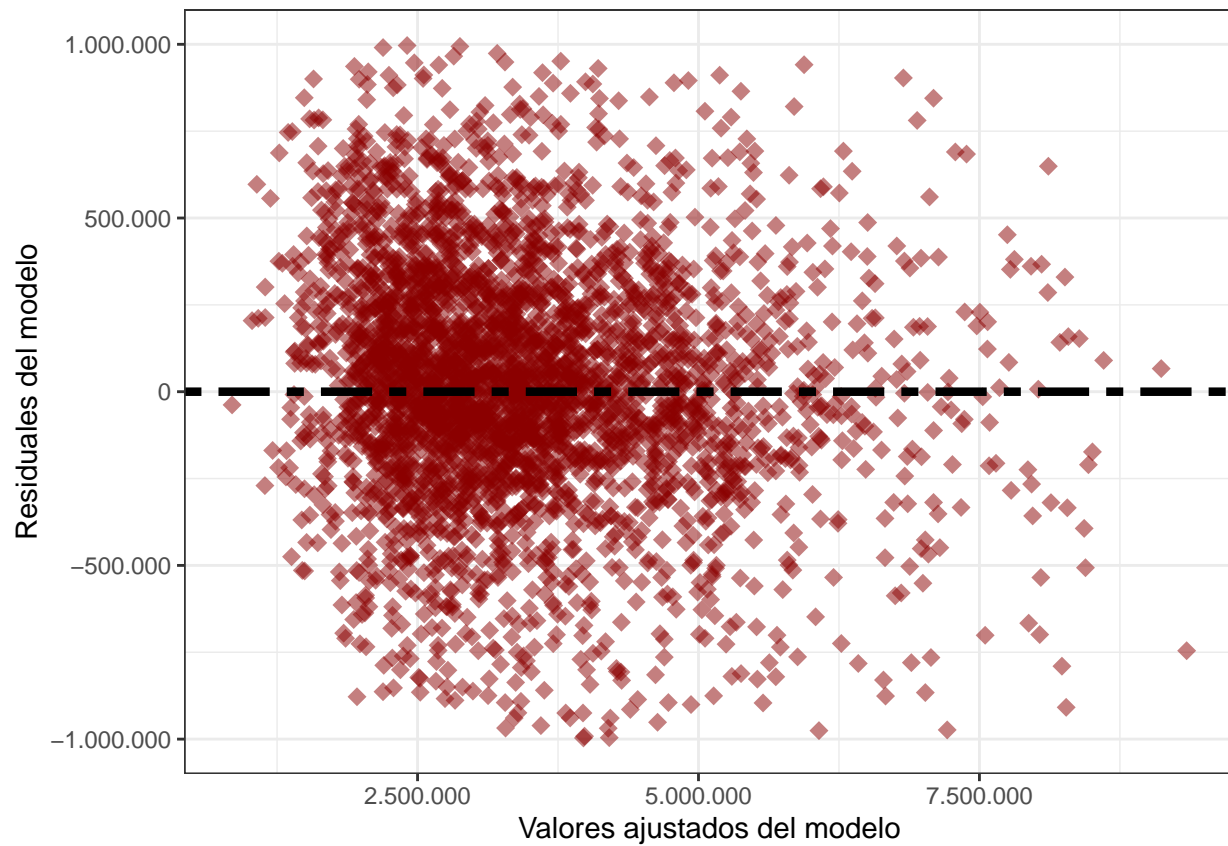
A	CODIGO_LOCALIDAD	TASA_RENTA	TASA_RENTA_2
2018	6	171.04	0.0058
2019	6	170.81	0.0059
2020	6	177.96	0.0056
2017	7	172.73	0.0058
2018	7	178.50	0.0056
2019	7	178.28	0.0056
2020	7	185.42	0.0054
2017	8	173.36	0.0058
2018	8	179.13	0.0056
2019	8	178.91	0.0056
2020	8	186.05	0.0054
2017	9	164.24	0.0061
2018	9	170.01	0.0059
2019	9	169.79	0.0059
2020	9	176.94	0.0057
2017	10	171.70	0.0058
2018	10	177.47	0.0056
2019	10	177.24	0.0056
2020	10	184.39	0.0054
2017	11	160.68	0.0062
2018	11	166.45	0.0060
2019	11	166.22	0.0060
2020	11	173.37	0.0058
2017	12	151.61	0.0066
2018	12	157.39	0.0064
2019	12	157.16	0.0064
2020	12	164.31	0.0061
2017	13	146.24	0.0068
2018	13	152.01	0.0066
2019	13	151.79	0.0066
2020	13	158.94	0.0063
2017	14	123.04	0.0081
2018	14	128.82	0.0078
2019	14	128.59	0.0078
2020	14	135.74	0.0074
2017	15	155.21	0.0064
2018	15	160.99	0.0062
2019	15	160.76	0.0062
2020	15	167.91	0.0060
2017	16	181.81	0.0055
2018	16	187.59	0.0053
2019	16	187.36	0.0053
2020	16	194.51	0.0051
2017	17	128.60	0.0078
2018	17	134.37	0.0074
2019	17	134.15	0.0075
2020	17	141.30	0.0071
2017	18	168.04	0.0060

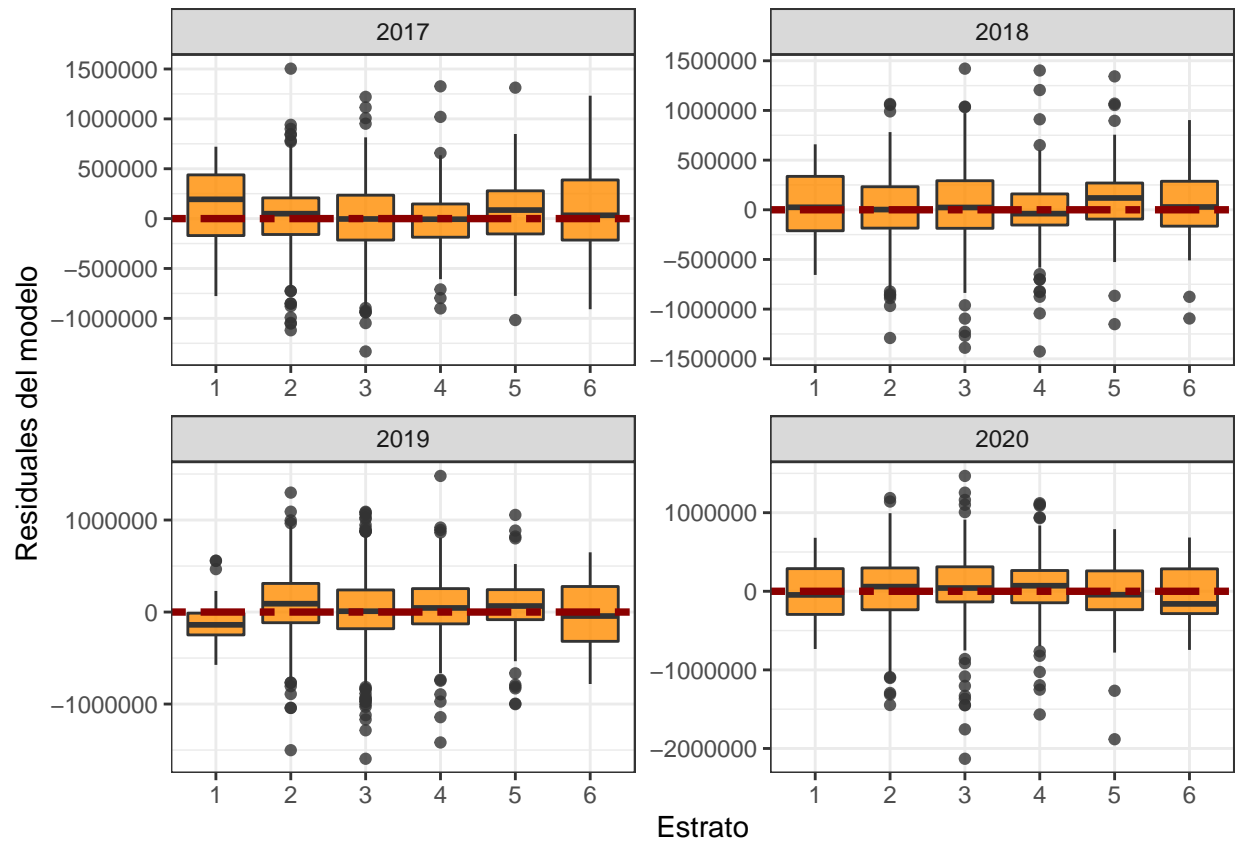
(continued)

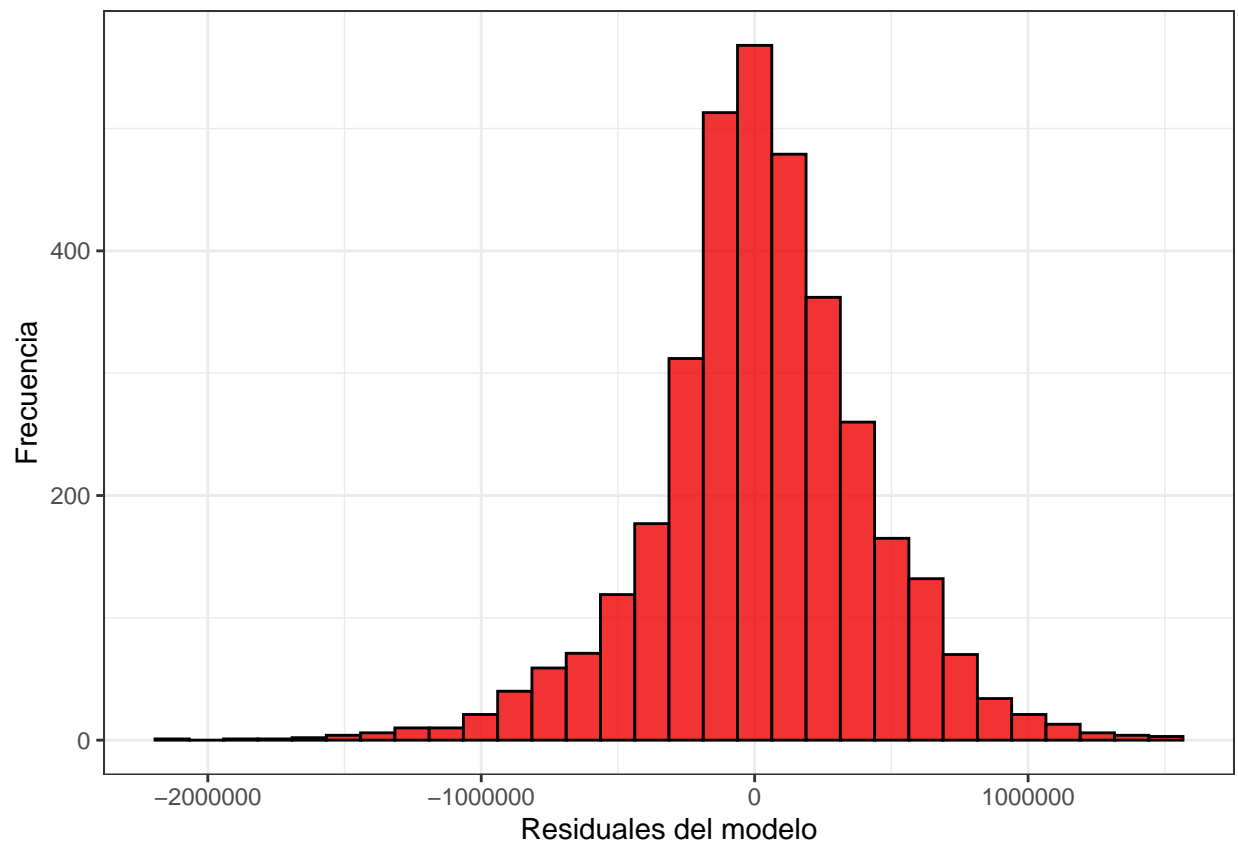
A	CODIGO_LOCALIDAD	TASA_RENTA	TASA_RENTA_2
2018	18	173.81	0.0058
2019	18	173.58	0.0058
2020	18	180.73	0.0055
2017	19	161.68	0.0062
2018	19	167.45	0.0060
2019	19	167.22	0.0060
2020	19	174.37	0.0057

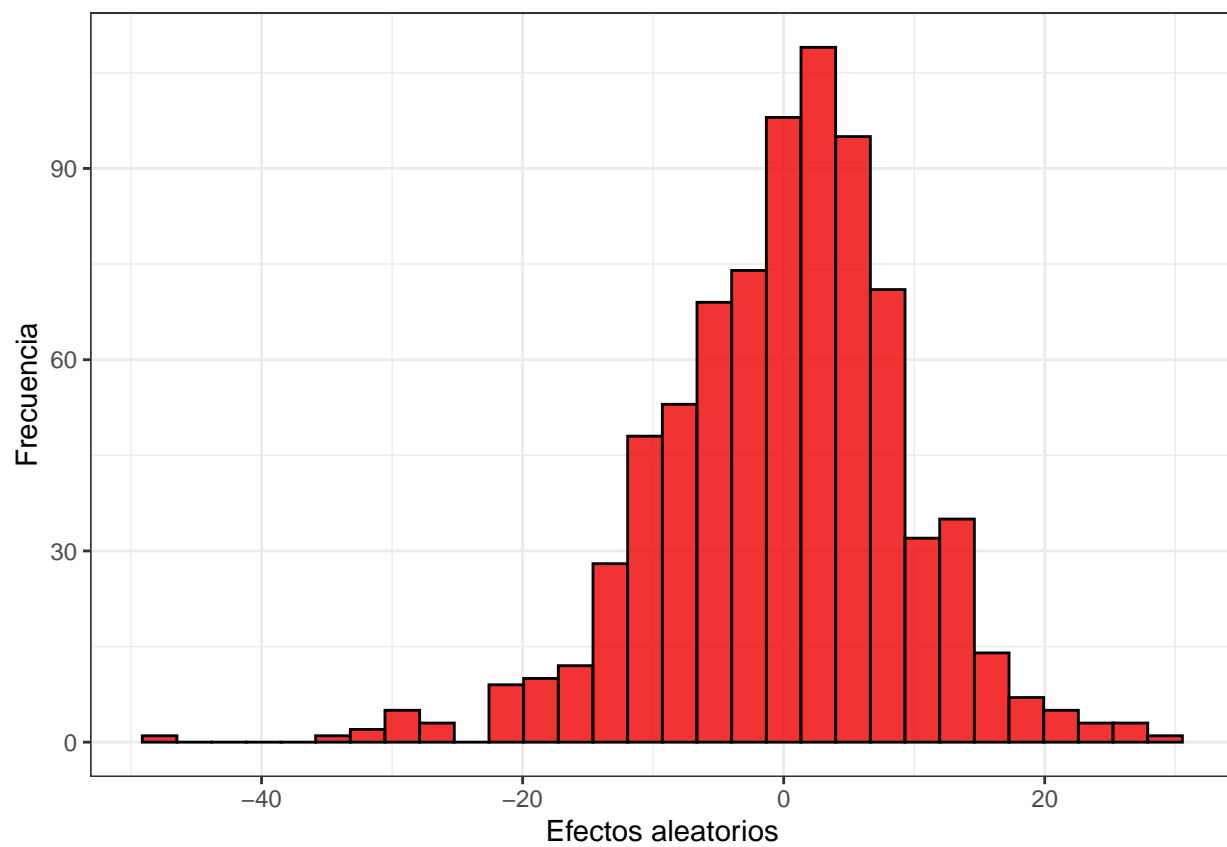
La base de datos consta originalmente con 1.390.326 registros. El resumen del número de registros, antes de realizar ciertas exclusiones se presenta en la siguiente tabla, la cual muestra el número de ofertas por fuente.

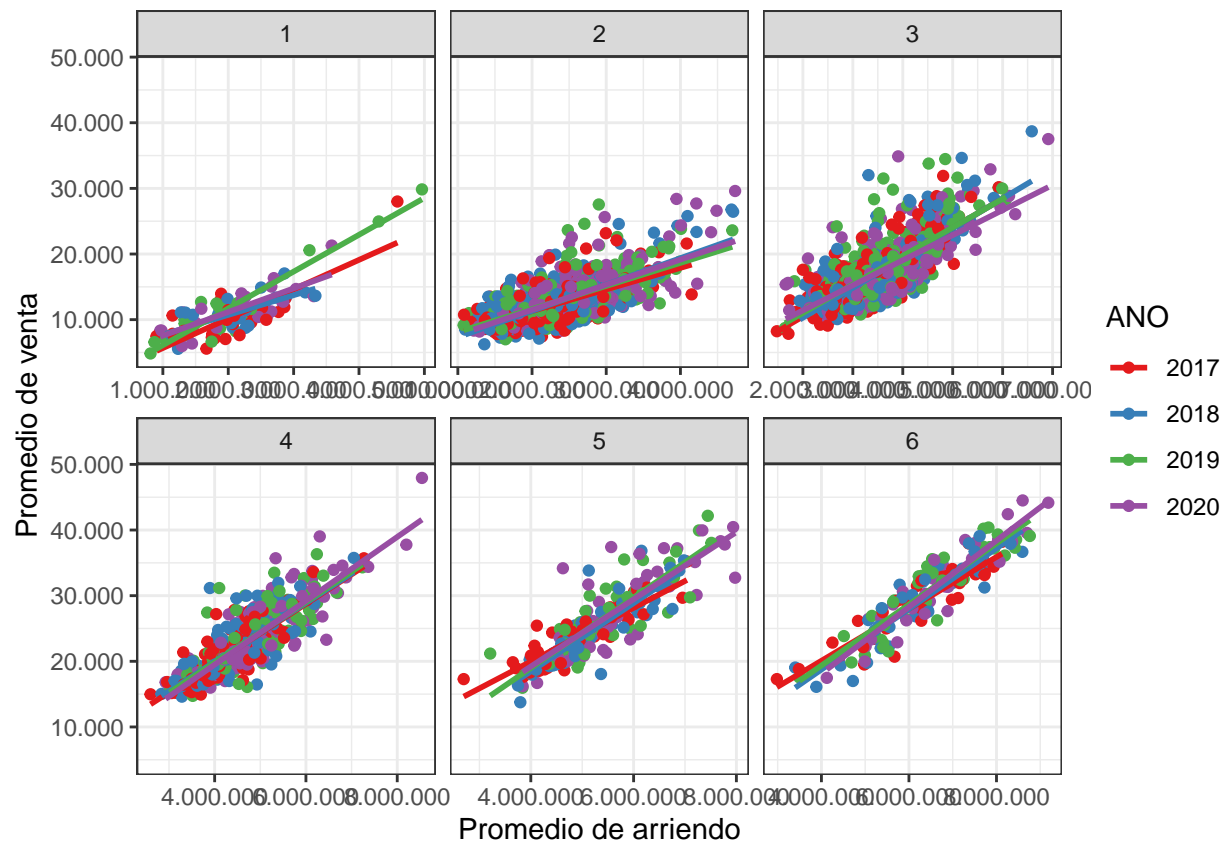
NOMBRE_FUENTE	N
FINCA RAIZ	1,112,410
GALERIA INMOBILIARIA	46,852
OIC	72,767
PROPERATI	158,297

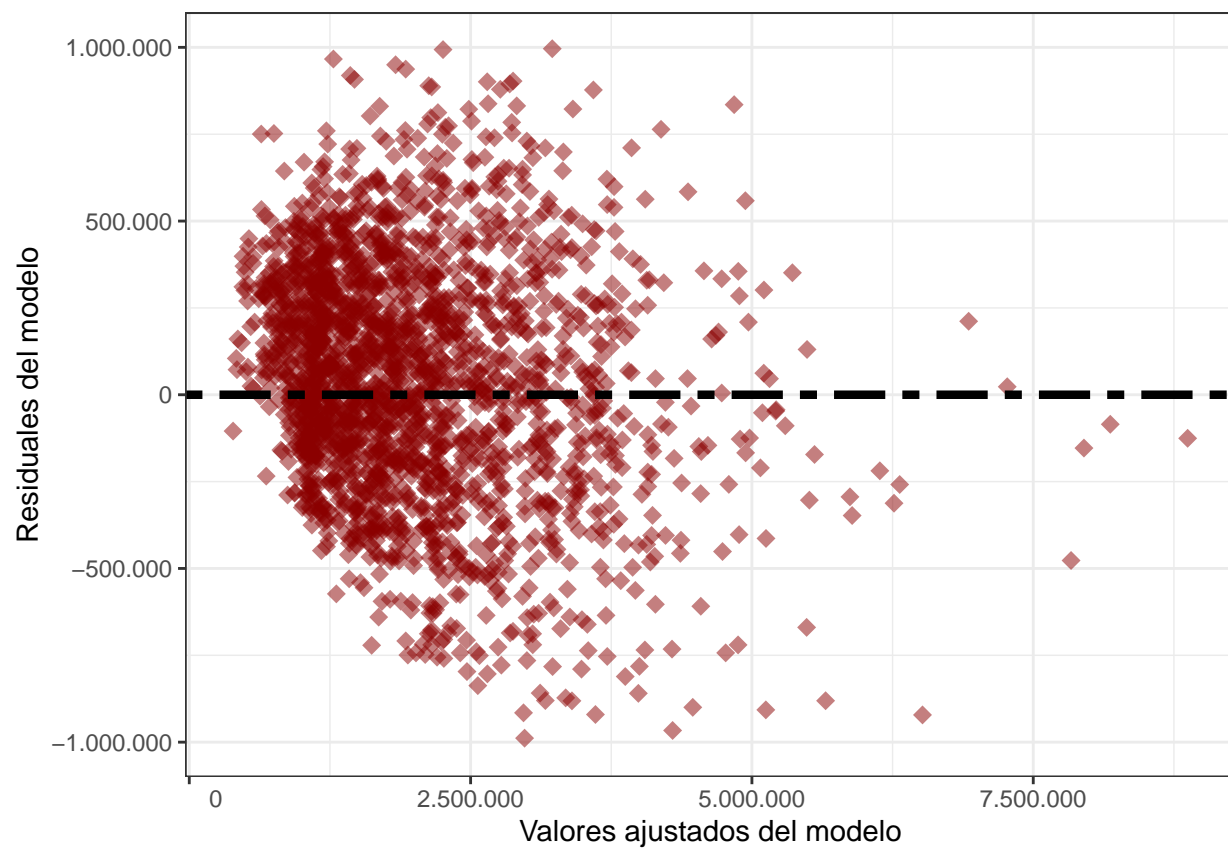


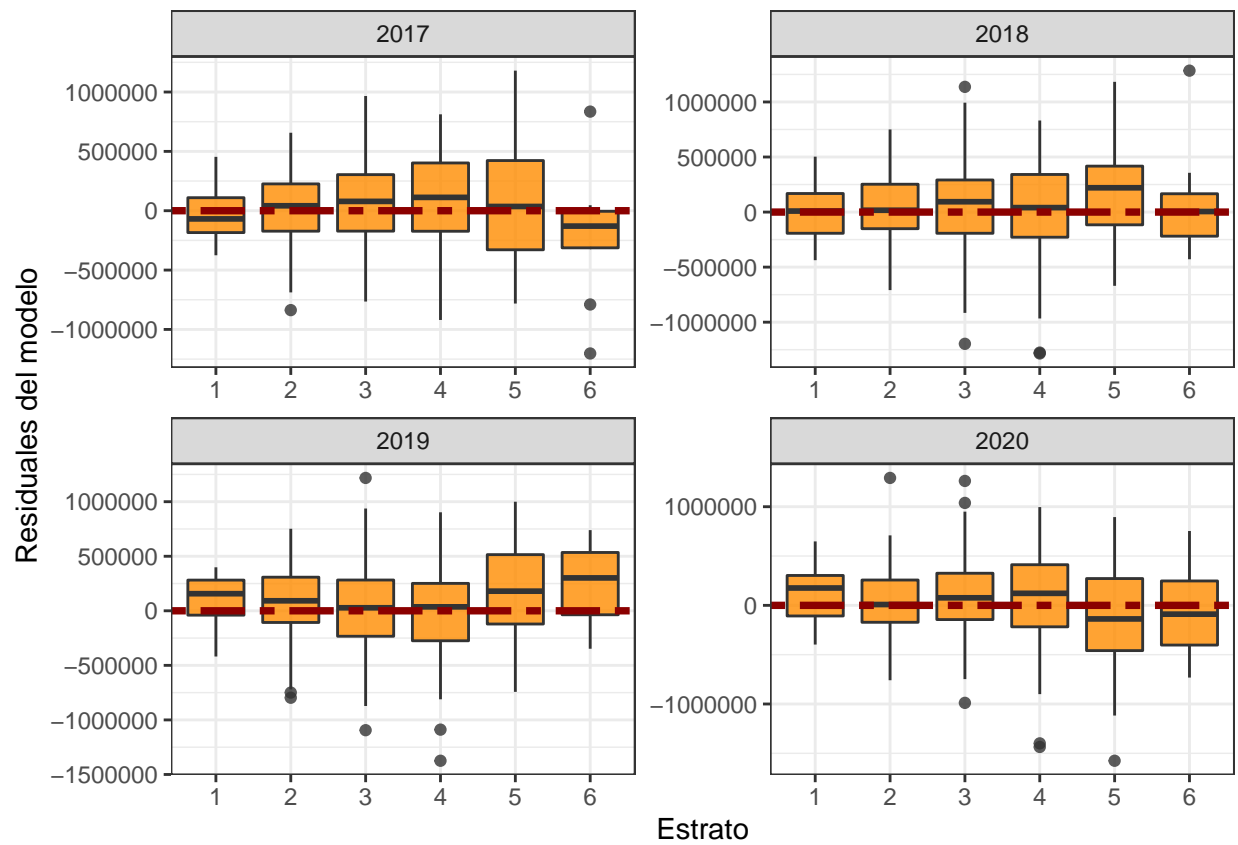


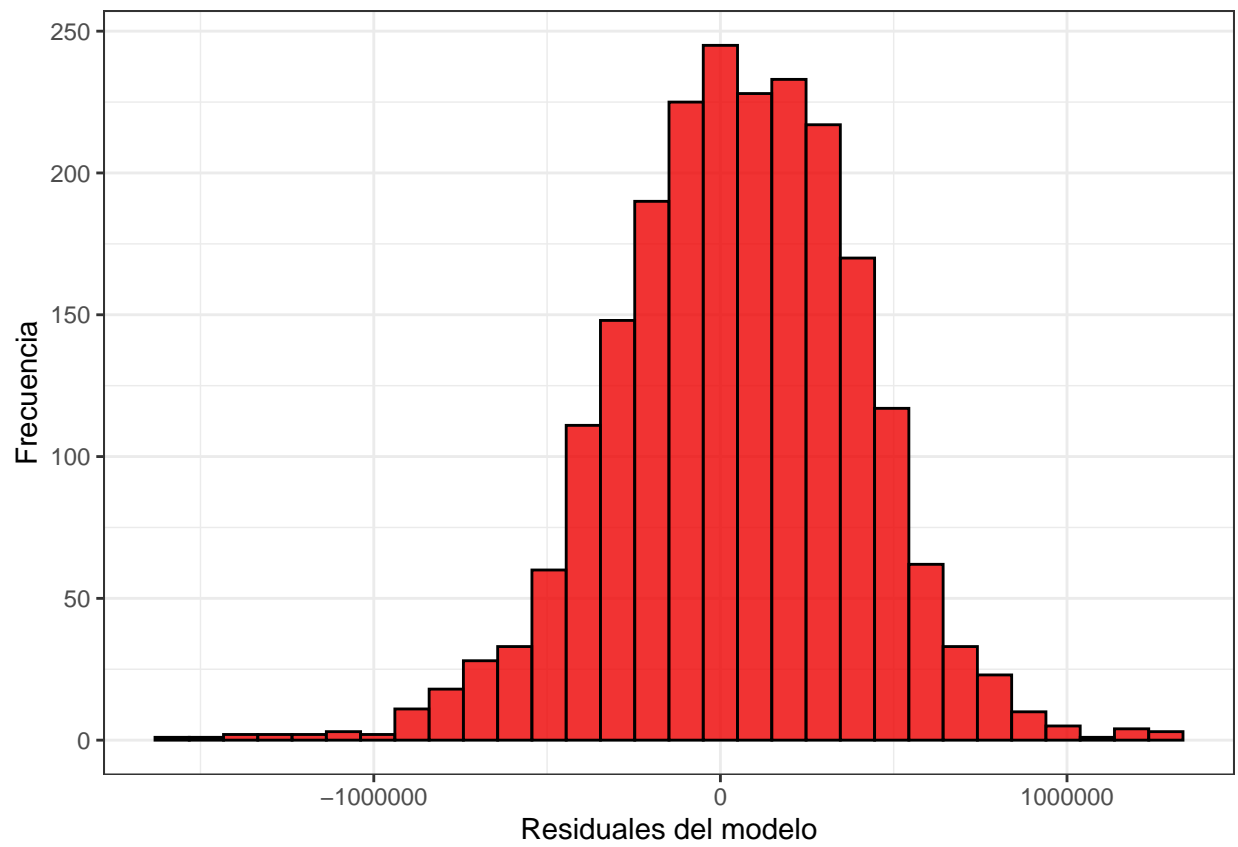


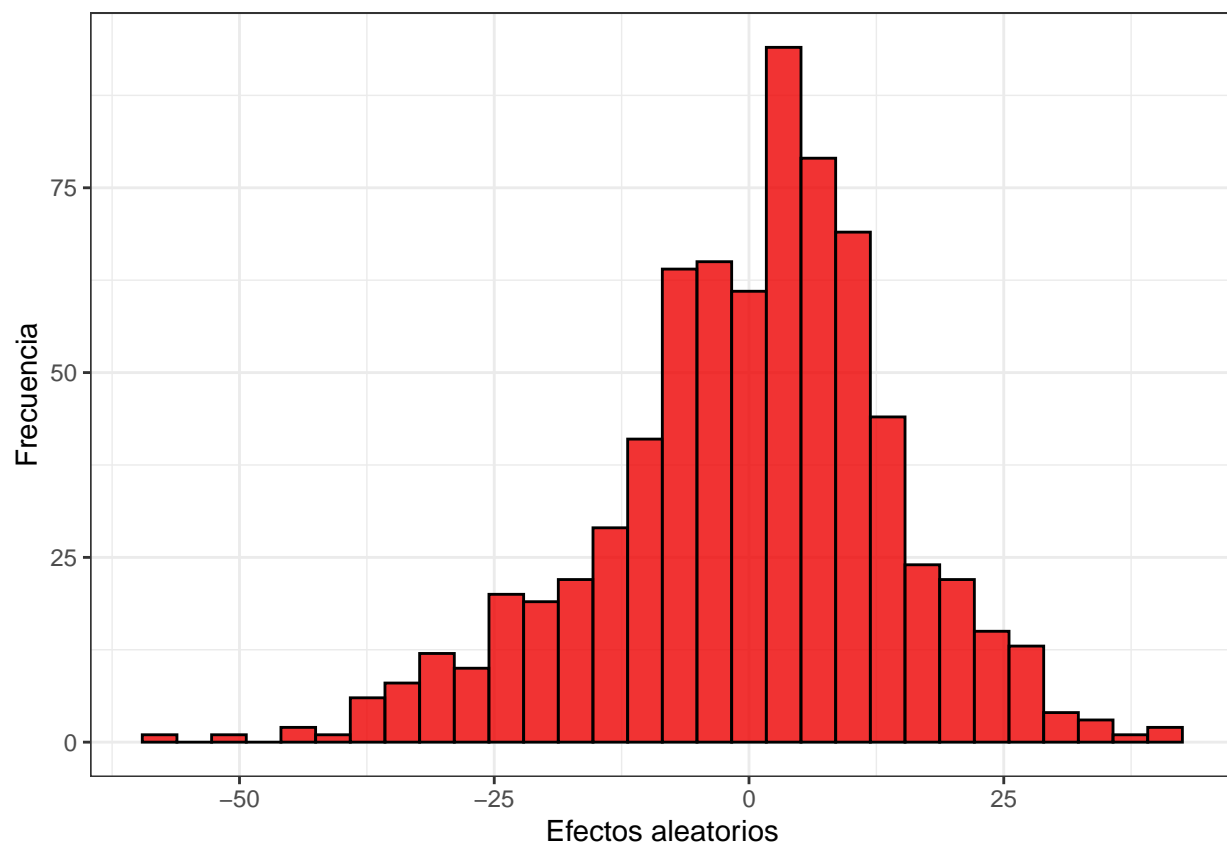


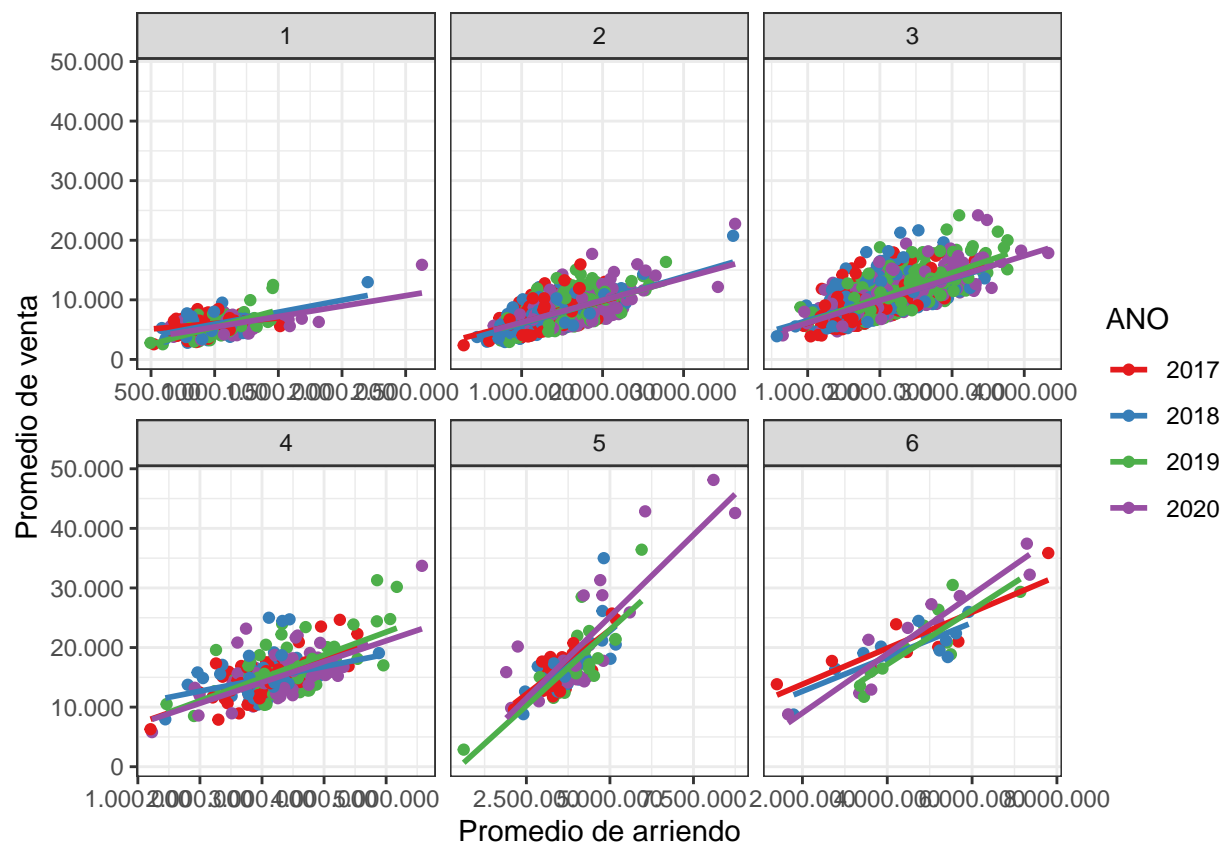




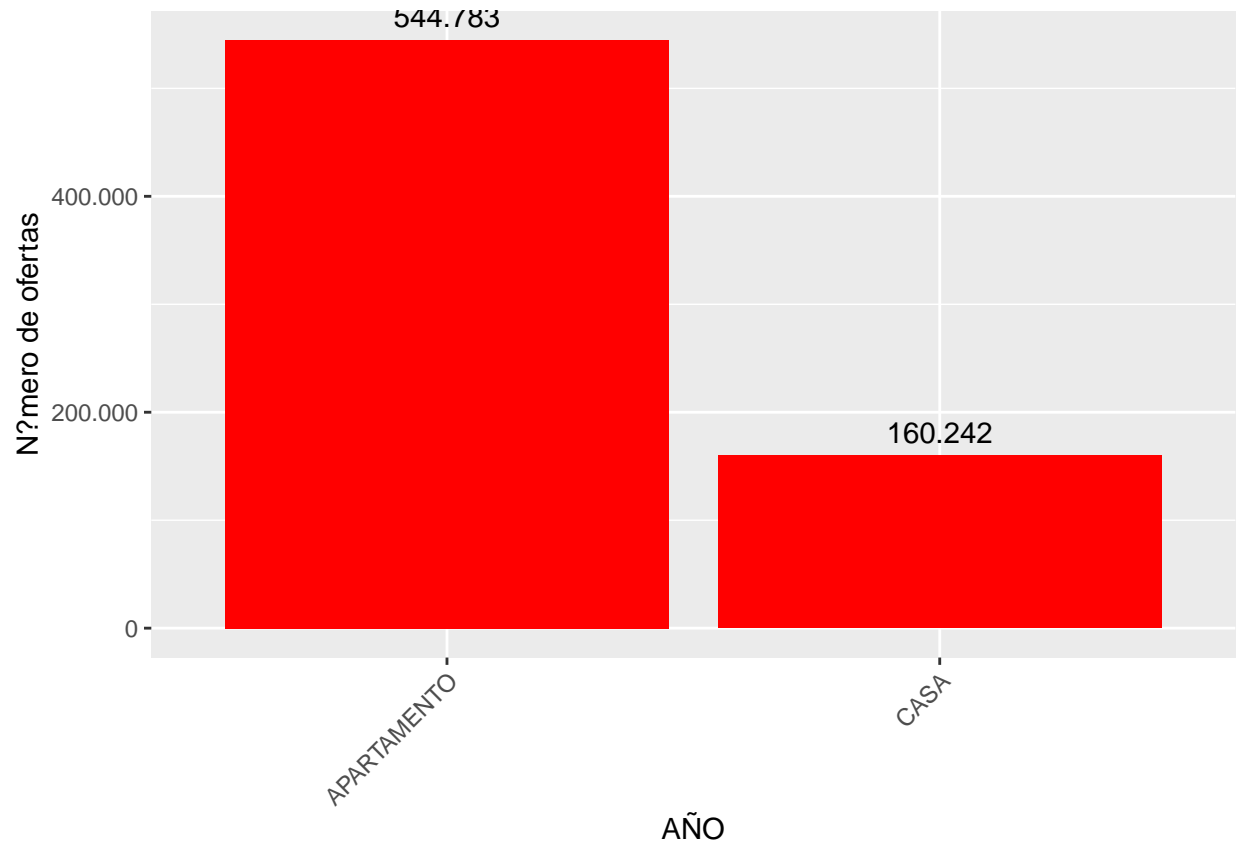


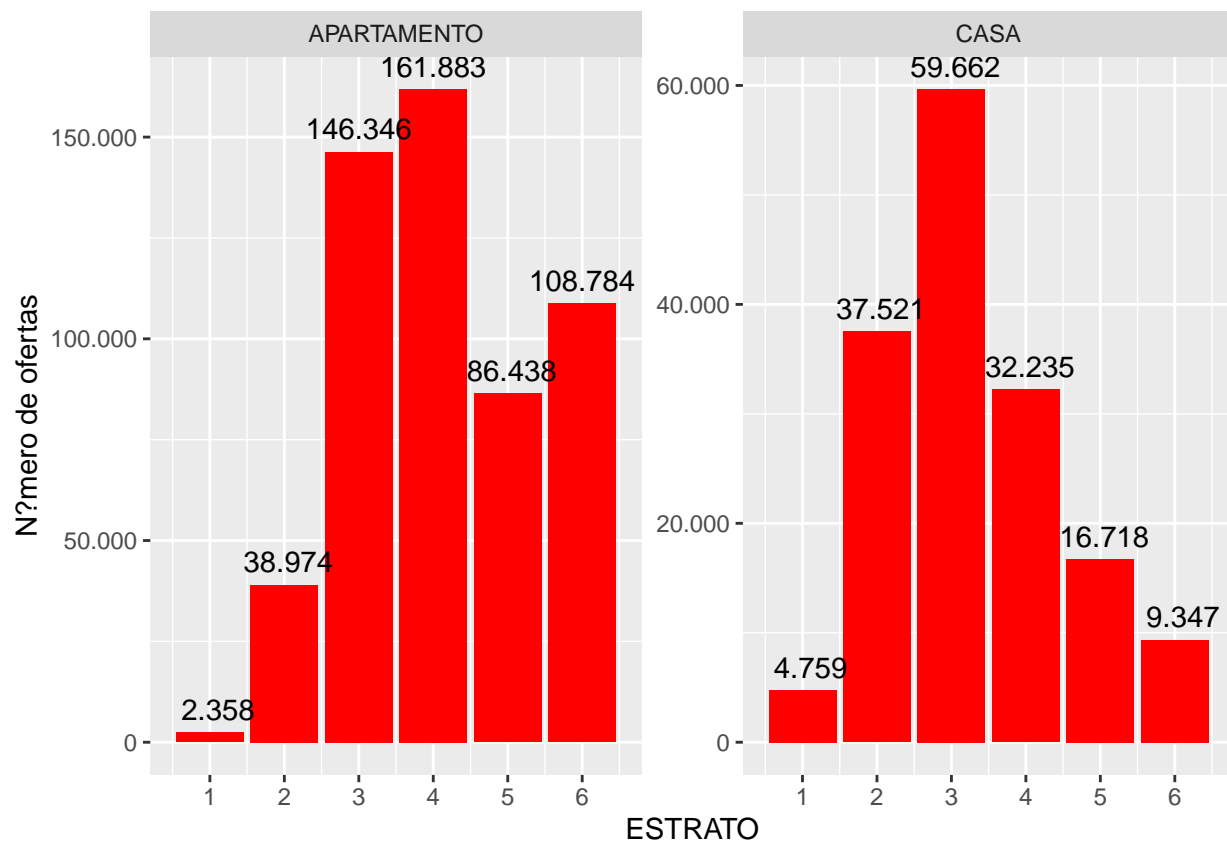


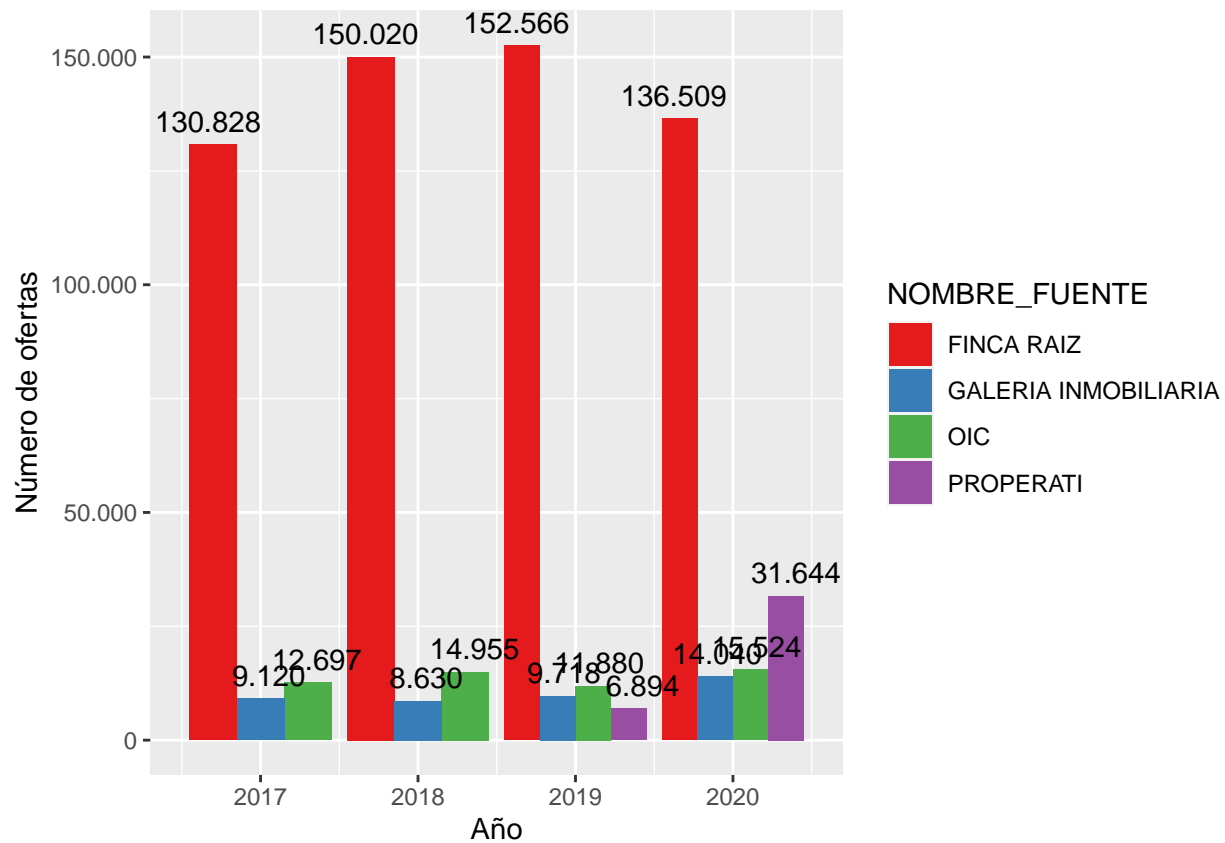


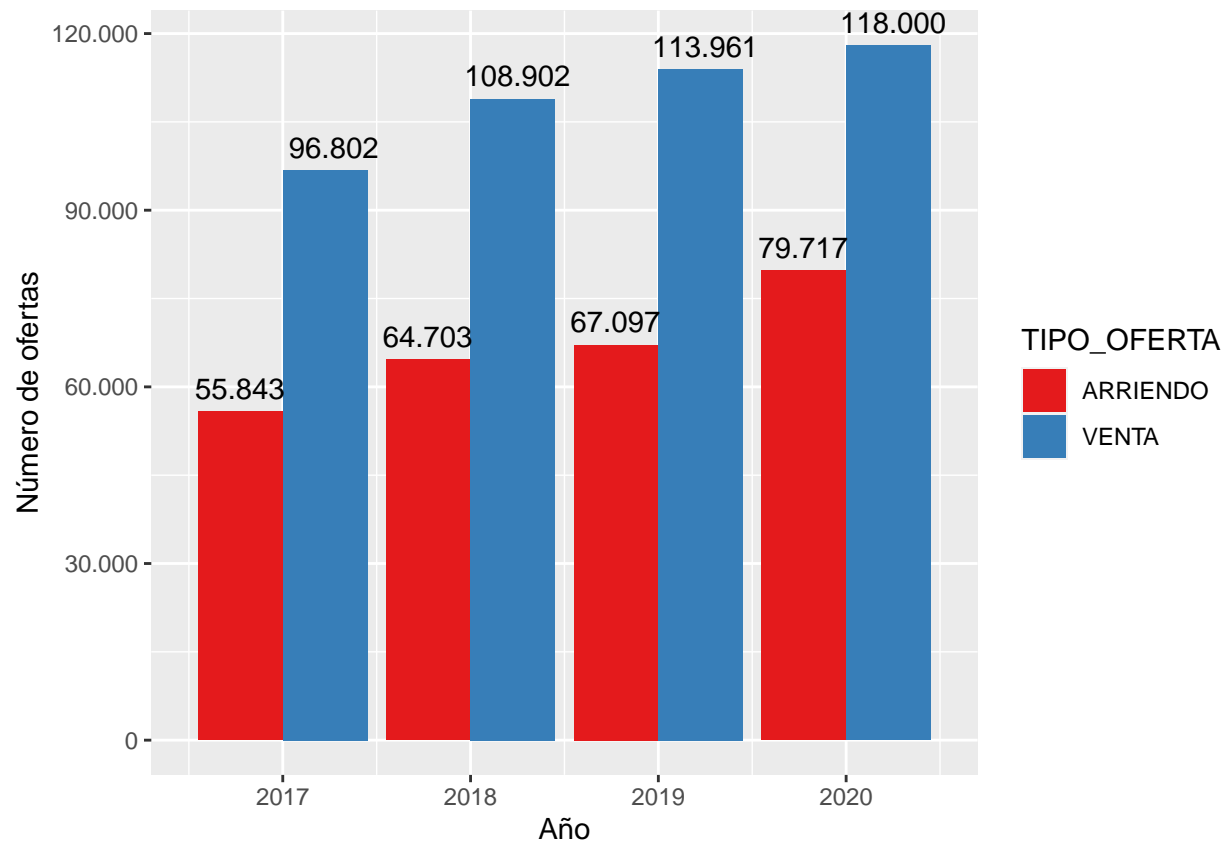


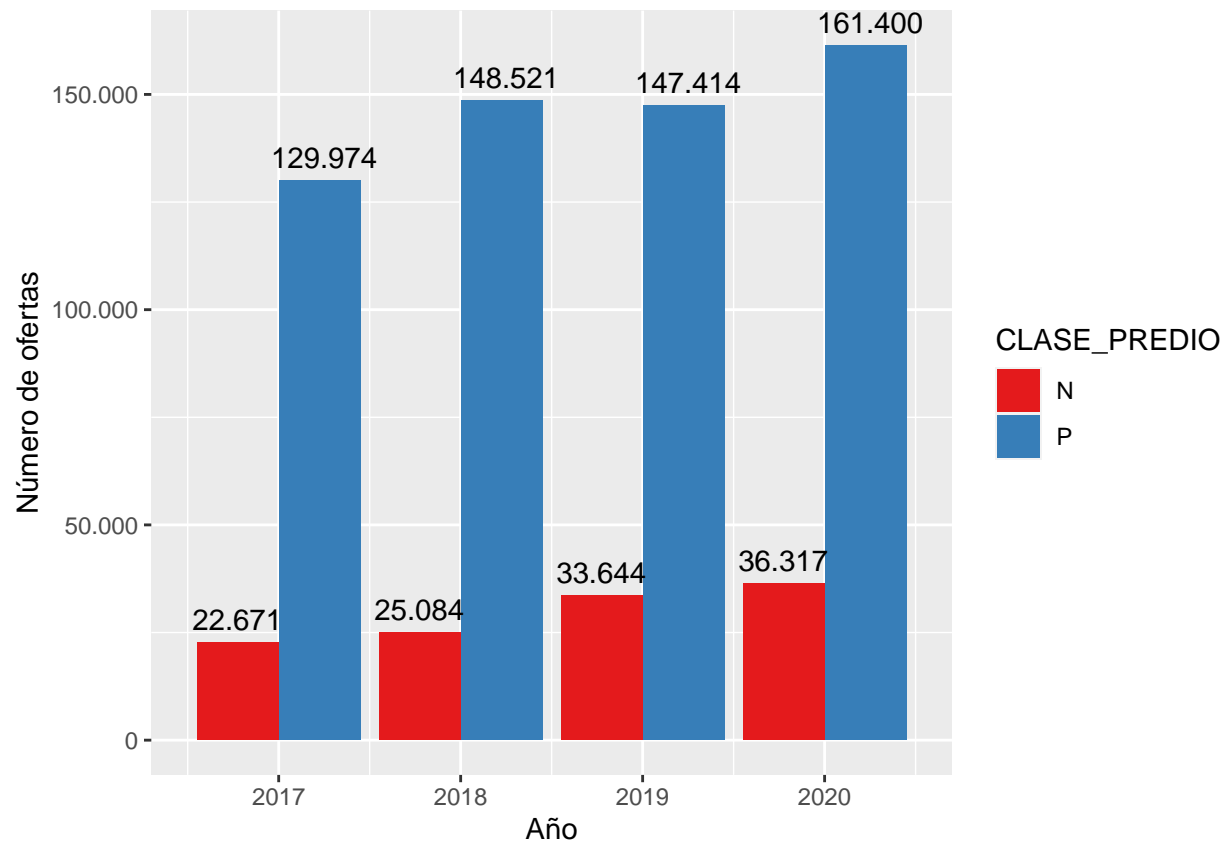
Inicialmente se filtraron aquellos registros que no cuentan con identificador, precio o barmanpre. Posterior a esta revisión se excluyeron los registros que no tenían área construida (2) y aquellos casos que no tenían área de terreno cuando el tipo de inmueble es diferente a apartamentos (2). A partir de esta base depurada se realizaron los siguientes conteos y gráficos de tipo descriptivo. Luego de estas exclusiones, se tiene un restante de 705.025 ofertas, las cuales son tenidas en cuenta dentro de los siguientes gráficos y tablas descriptivas.

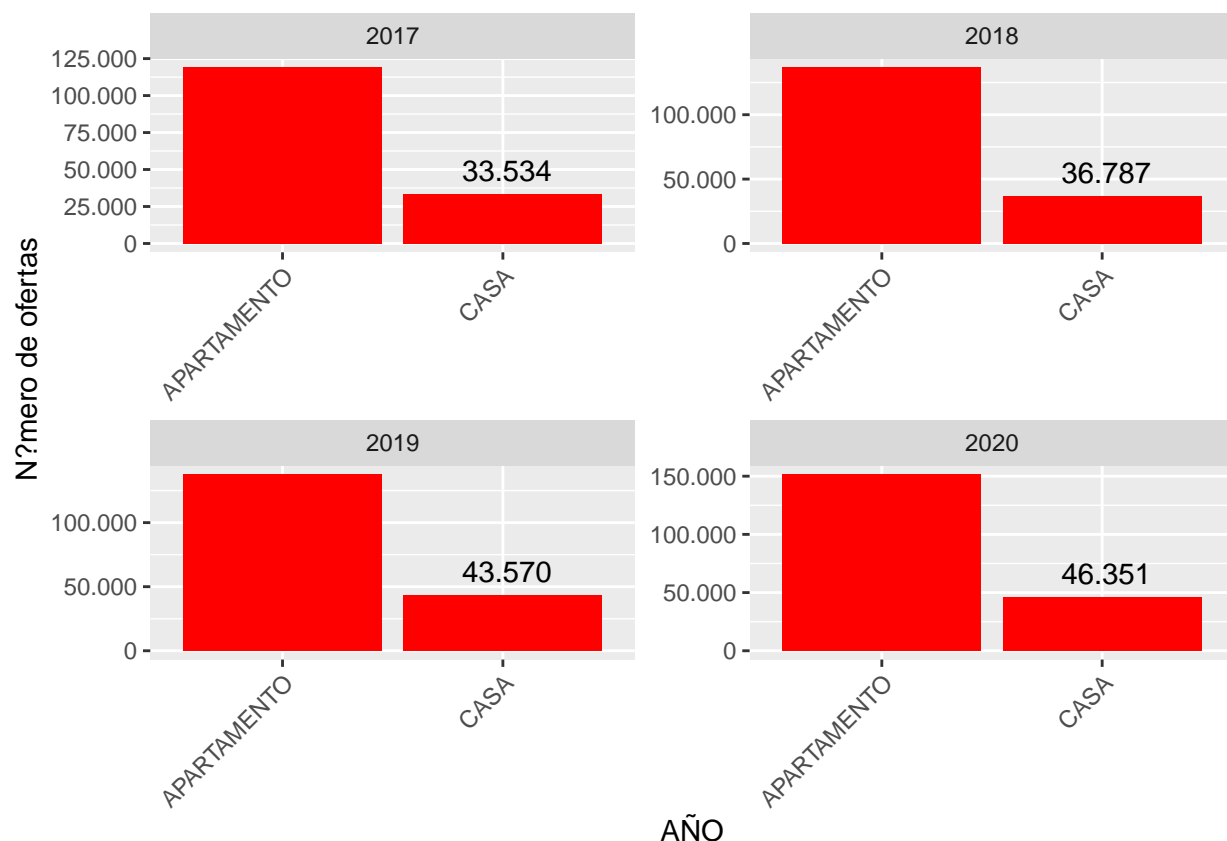












References

- 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.
- O Melo, L López, and S Melo. Diseño de experimentos: métodos y aplicaciones. *Editorial Universidad Nacional de Colombia. Bogotá*, 2007.
- José Pinheiro and Douglas Bates. *Mixed-effects models in S and S-PLUS*. Springer Science & Business Media, New York, 2006.
- Oliver Schabenberger and Carol A Gotway. *Statistical methods for spatial data analysis*. Chapman and Hall/CRC, Boca Raton, 2017.
- 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.