PREDICCIÓN DEL VALOR DE UN INMUEBLE MEDIANTE TÉCNICAS AGREGATIVAS.

Juan José Goyeneche [1]
Departamento de Métodos Cuantitativos, Iesta, Fcea, Udelar
Leonardo Moreno [2]
Departamento de Métodos Cuantitativos, Iesta, Fcea, Udelar.
Marco Scavino [3]
Departamento de Métodos Cuantitativos, Iesta, Fcea, Udelar.

RESUMEN

El objetivo del trabajo es construir un modelo que permita efectuar de manera eficiente, en términos predictivos, la estimación del valor contado de un determinado inmueble. Se tiene como insumo la información espacio-temporal sobre el valor contado de un conjunto de inmuebles y ciertas variables hedónicas intrínsecas a cada bien, por ejemplo, superficie, antigüedad, número de dormitorios. En tal sentido es conocido el valor de ciertas propiedades en diferentes fechas, entendiendo por valor contado aquel que es asignado por el tasador, donde cada propiedad puede haber sido tasada en diferentes momentos, problema denominado en la literatura como ventas repetidas. La información espacial permite la construcción de modelos autorregresivos espaciales donde el precio de un inmueble se encuentra correlacionado con el de sus vecinos. La información temporal es modelada mediante un modelo autorregresivo temporal, en este caso la dependencia entre una tasación y la siguiente disminuye en función al tiempo transcurrido entre una y otra. Las variables hedónicas se modelan mediante un modelo de regresión lineal dinámico, donde los coeficientes de la regresión son función de la ubicación del inmueble. A partir de la metodología de agregación de Stacking se busca un procedimiento que permita predicciones más precisas. En este caso, otras metodologías son incluídas en la agregación. Se evalúa la performance mediante el error cuadrático en una muestra de testeo.

Palabras clave: *Método de Stacking, modelos espacio-temporales, precio hedónico, procesos autorregresivos.*

Área de conocimiento: Métodos Matemático-Cuantitativos

[1] *email:*`jjgoye@iesta.edu.uy`
[2] *email:*`mrleo@iesta.edu.uy`
[3] *email:*`mscavino@iesta.edu.uy`

1 Introducción

Es de vital importancia para determinadas políticas gubernamentales o entidades financieras conocer como se comporta de forma temporal y espacial los precios de ciertos inmuebles, según sus características, en una determinada ciudad. En tal camino nuestro trabajo apunta a enmarcar estos conceptos en la ciudad de Montevideo, Uruguay. Cabe aclarar que este trabajo no sustituye el trabajo del tasador, como es sabido la tasación de un inmueble es una arte que depende de aspectos muchas veces no cuantificables. Si embargo consideramos que los resultados obtenidos permiten determinar ciertos patrones en cuanto al precio de las propiedades, lo cual puede ser de vital importancia en la toma de decisiones a nivel político o empresarial.

En la últimas décadas se han elaborado distintas metodologías para poder predecir el valor actual de un inmueble. Dichos trabajos basan sus modelos en distintos tipos de información, la cual podemos clasificar en tres grupos:

1. (Información espacial) Estos modelos basan la predicción del valor del inmueble en función del precio de sus inmuebles vecinos. Para definir vecinos es necesario establecer una distancia, la cuál puede incluir aspectos geográficos y/o económicos. De tal forma se puede establecer un nivel de influencia, un peso en el precio del inmueble j, sobre el inmueble i, anotaremos w_{ij}. Si llamamos Y_i al precio del inmueble i entonces,

$$Y_i = \sum_{j \neq i} w_{ij} Y_j \tag{1}$$

   En general se impone que $w_{ij} \geq 0 \, \forall i, j$ y $\sum_j w_{ij} = 1 \, \forall i$.

2. (Información hedónica) Esta información es aquella intrínseca al inmueble, como por ejemplo la superficie, número de dormitorios, número de baños, la antigüedad de la construcción y otro gran conjunto de covariables que caracterizan al bien, anotaremos $X_1, \ldots, X_p$ a dichas covariables. Los modelos hedónicos parten del supuesto que el precio del inmueble puede ser descompuesto como suma del precio de sus características o atributos, es decir se trabaja con un modelo de regresión lineal donde las covariables son los atributos y el output es el precio del inmueble,

$$Y_i = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_p X_{ip} + \varepsilon_i, \tag{2}$$

   siendo $\beta_0, \ldots, \beta_p$ los coeficientes del modelo y ε_i los errores que se asumen en general i.i.d. con distribución normal centrada y varianza constante. Los modelos llamados económetricos espaciales, ver [Anselin, 1988], ingresan en la estimación ambos puntos anteriores, es decir,

$$Y_i = \alpha \sum_{j \neq i} w_{ij} Y_j + \beta_0 + \beta_1 X_{i1} + \ldots + \beta_p X_{ip} + \varepsilon_i, \tag{3}$$

   donde α es llamado el parámetro de referencia, que regula la importancia del precio de los vecinos en la modelización. En el trabajo emplearemos esta metodología con el agregado, al igual que en [Sun et al., 2014], que consideraremos los coeficientes del modelo como función de la posición, es decir, se estiman funciones que dependen de la ubicación del inmueble.

3. (Información temporal) En este caso es conocido el precio del inmueble pero en uno o en varios momentos temporales pasados, y es necesario actualizar dicha información. Se ha comprobado que estos modelos, llamados modelos de ventas repetidas, son muchos más eficaces para predecir el precio actual de un inmueble, y donde es necesario estimar el índice de precios. Son extensos los trabajos referentes al tema, [Nagaraja et al., 2010], [Bailey et al., 1963], [Shiller and Case, 1987], en estos casos, mediante diferentes metodologías se modela el logaritmo del precio del inmueble. Recientemente se han propuestos modelos autorregresivos, [Nagaraja et al., 2011], que permiten introducir en el modelo alguna característica propia del inmueble como por ejemplo el efecto del barrio donde se encuentra.

Sin embargo son escasos los trabajos que utilizan simultáneamente toda la información mencionada. En tal sentido, a través de la metodología de agregación llamada Stacking, desarrollado por Leo Breiman, [Breiman, 1996], el objetivo del trabajo es implementar algunas de estas técnicas descritas, obtener distintas predicciones del inmueble, y mediante la agregación generar una única predicción. Seleccionando una muestra de testeo se evaluará la mejora en las predicciones en términos del error cuadrático medio.

En la sección 2 siguiente se describen brevemente las metodologías mencionadas. En la sección 3 se realiza un análisis descriptivo de la base de datos. Debido a una cierta cantidad de datos faltantes y datos erróneos es necesario eliminar algunos inmuebles de la base. En la sección 4 se expresan los valores predichos por cada modelo y los errores producidos. Se observa como en este caso el método de agregación es más eficiente en términos de errores de predicción. En la sección 5 se establecen las conclusiones y trabajos futuros.

2 Marco teórico

En esta sección se realiza una breve reseña de diversos métodos de la literatura para predecir el valor de un inmueble a partir de diferente información. Por último se desarrolla el método de agregación de Stacking.

2.1 Modelo Semiparamétrico Espacial Dinámico

El modelo tiene por insumos las características hedónicas del inmueble (ej: superficie, número de dormitorios) y la ubicación espacial del inmueble, ver [Sun et al., 2014]. La idea consiste en modelar el precio de la propiedad i-ésima, considerando la dependencia espacial con los inmuebles vecinos de manera similar a los modelos ecónometricos espaciales, [Anselin, 1988], y por otro lado un modelo lineal que depende de covariables intrínsecas al inmueble donde los coeficientes del modelo no son constantes, sino funciones que dependen de las coordenadas del inmueble. El modelo propuesto es el siguiente,

$$Y_i = \alpha \sum_{j \neq i} w_{ij} Y_j + X_i^T \beta(s_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{4}$$

donde,

- X_i es el vector $p-$dimensional de covariables asociadas al inmueble con valor Y_i.
- s_i es un vector bidimensional que indica la posición del inmueble i.
- w_{ij} son pesos conocidos que miden una distancia ecónomica entre el inmueble i y el inmueble j, en nuestro trabajo dicha distancia sólo dependerá de la ubicación geográfica de los inmuebles.

- $\beta(\cdot) = (\beta_1(\cdot), \ldots, \beta_p(\cdot))^T$ es un vector cuyas componentes son función de la posición.
- Los errores del modelos ε_i son considerados con distribuciones normales centradas con varianza σ^2, independientes del vector de covariables.
- α es un parámetro desconocido llamado parámetro de referencia que regula el efecto de la dependencia de los vecinos del inmueble.

En este modelo se tomo como la matriz de pesos,

$$w_{ij} = \frac{e^{\|s_i - s_j\|}}{\sum_{k \neq j} e^{\|s_i - s_k\|}} \tag{5}$$

donde s_i es el vector bivariado que indica la coordenadas del inmueble i-ésimo de la base.

A modo de ejemplo se consideran como covariables, el número de dormitorios, la superficie total y la antigüedad de la propiedad. El modelo no incorpora tasaciones pasadas.

2.2 Estimación no-paramétrica del índice temporal para ventas repetidas

En este modelo se cuenta en la muestra de entrenamiento con inmuebles que poseen un valor de venta “vigente”, es decir un valor de venta estipulado a partir de una determinada fecha, y al menos un valor de venta “no-vigente” lo que significa que tuvo al menos un valor de venta asignado anterior a una determinada fecha. El modelo tiene por objetivo predecir el valor de venta vigente a través del valor de venta no-vigente. Para ello se realiza la estimación de una función temporal del índice de precios.

Se establece entonces que la relación de tasaciones entre un momento pasado, t, y una tasación actual en el momento t_0, depende de la distancia entre t_0 y t, esto es, del momento en que se hizo la tasación histórica y el momento en que se hace la tasación actual. La variable a ser modelada es entonces,

$$\text{Indice al tiempo } t = \frac{\text{tasación en } t_0}{\text{tasación en } t}.$$

Al índice al tiempo t lo anotaremos I_t. Los inmuebles de la muestra de entrenamiento los llamaremos *inmuebles del modelo*. El número de relaciones posibles entre los inmuebles del modelo serán llamadas *puntos del modelo*. Cabe aclarar que no es contradictorio que el número de puntos del modelo sea superior al número de inmuebles del modelo, puesto que a una tasación vigente le puede corresponder más de una tasación no vigente del mismo inmueble. Por algún método no paramétrico se estima la función índice para todo t, en nuestro caso por splines cúbicos. Dicha función nos permitirá actualizar las tasaciones no-vigentes de aquellos inmuebles que no tengan tasación vigente. Este modelo sólo incorpora información de ventas anteriores. Por tanto inmuebles que no tengan alguna tasación realizada no pueden ser evaluados en este modelo.

2.3 Modelo de efectos mixtos

Unos modelos sencillos, pero plausibles para la modelización de este tipo de problemas son los modelos mixtos, ver [McCulloch and Searle, 2001]. Podemos tratar al efecto del tiempo, β_t, como un efecto fijo, el efecto inmueble (α_i) y el barrio de la propiedad (τ_z) como efectos aleatorios. Dicha metodología no tiene una serie temporal como componente en el modelo, es decir, no toma en cuenta el orden de los años. La información espacial es débilmente incluída en el efecto barrio. Se expresa de la siguiente manera,

$$y_{i,j,z} = \mu + \alpha_i + \tau_z + \beta_{t(i,j,z)} + \varepsilon_{(i,j,z)}, \tag{6}$$

donde,

- $\alpha_i \overset{\text{i.i.d}}{\sim} N(0, \sigma^2_\alpha)$
- $\tau_z \overset{\text{i.i.d}}{\sim} N(0, \sigma^2_\tau)$
- $\varepsilon_{(i,j,z)} \overset{\text{i.i.d}}{\sim} N(0, \sigma^2_\varepsilon)$
- μ es un parámetro fijo y desconocido.
- Los barrios z toma valores 1 a Z, los inmuebles i en cada barrio van de 1 a I_z, y j-ésima venta del inmueble i toma valores de 1 a J_i.

2.4 Modelos de B-M-N y C-S

Bailey, Muth y Nourse, [Bailey et al., 1963], introducen un modelo, anotaremos B-M-N, en referencia la logaritmo del valor del inmueble, es decir, si partimos de la siguiente expresión,

$$\frac{Y_{it}}{Y_{it'}} = \frac{I_t}{I_{t'}} U_{itt'}, \tag{7}$$

siendo Y_{it} el precio del inmueble i en el tiempo t, I_t el índice de precios al tiempo t y $U_{tt'}$ un factor aleatorio. Podemos suponer que el error $U_{tt'}$ tiene distribución LogNormal,

$$u_{itt'} = \log U_{itt'} \overset{i.i.d.}{\sim} N(0, \sigma^2_u).$$

Si aplicamos logaritmo a la ecuación (7) obtenemos,

$$y_{it} - y_{it'} = b_t - b_{t'} + u_{itt'}, \tag{8}$$

con $y_{it} = \log Y_{it}$ y $b_t = \log I_t$. Obtenemos así un modelo lineal para el tiempo entre ventas. Este modelo no tiene en cuenta en la regresión el efecto del lapso de tiempo entre una tasación y la siguiente.

El modelo de Case y Shiller, anotamos C-S, [Shiller and Case, 1987], al igual que B-M-N afronta el problema de ventas repetidas, pero con la idea de ponderar de forma distinta las observaciones en la ecuación (8) según el tiempo entre las tasaciones. Al igual que en los modelos anteriores se anota $y_{i,t}$ al logaritmo del precio del inmueble i en el tiempo t. El modelo es

$$y_{i,t} = \beta_t + H_{i,t} + \varepsilon_{i,t} \tag{9}$$

- β_t es es logaritmo del índice de precios en el tiempo t
- $\varepsilon_{i,t}$ son variables i.i.d con distribución normal con media cero y varianza σ^2.
- $H_{i,t}$ es un paseo al azar gausianno que incorpora la información acerca de las ventas previas.

Para su implementación ver [Nagaraja et al., 2010].

2.5 Modelo autorregresivo para ventas repetidas

Modelo introducido en [Nagaraja et al., 2011]. Se anota $y_{i,j,z}$ al logaritmo del precio de la j-ésima venta de la propiedad i, que se encuentra ubicada en el barrio z. Se considera al tiempo en períodos discretos $1,2,\ldots,T$, donde $t(i,j,z)$ es el tiempo en que ocurre la j-ésima venta de la propiedad i, que se encuentra ubicada en el barrio z. Los lapsos de tiempo entre ventas consecutivas son

$$\gamma(i,j,z) = t(i,j,z) - t(i,j-1,z)$$

Finalmente, el total de observaciones de la muestra de entrenamiento es $N = \sum_{z=1}^{Z}\sum_{i=1}^{I_z} J_i$ donde Z es el total de barrios, I_z el número total de inmuebles en el barrio z, y J_i el número de ventas del inmueble i. El logaritmo del precio de venta es modelado mediante un modelo autorregresivo,

$$y_{i,1,z} = \mu + \beta_{t(i,1,z)} + \tau_z + \varepsilon_{i,1,z} \quad \text{si} \quad j = 1, \tag{10}$$

$$y_{i,j,z} = \mu + \beta_{t(i,j,z)} + \tau_z + \phi^{\gamma(i,j,z)}\left(y_{i,j-1,z} - \mu - \beta_{t(i,j-1,z)} - \tau_z\right) + \varepsilon_{i,1,z} \quad \text{si} \quad j > 1, \tag{11}$$

donde,

- $\beta_{t(i,j,z)}$ es es logaritmo del índice de precios en el tiempo $t(i,j,z)$. Son asumidos entonces los logaritmos de los índices de precios $\beta_1,\ldots,\beta_T$ como efectos fijos.
- ϕ es un coeficiente autorregresivo con $|\phi| < 1$
- τ_z con $z = 1,\ldots,Z$ son el código barrio de la propiedad, que es considerado un efecto aleatorio. En este caso $\tau_z \overset{\text{i.i.d}}{\sim} N(0,\sigma_\tau^2)$.
- Se impone la restricción que $\sum_{t=1}^{T} n_t\beta_t = 0$ siendo n_t el número de ventas al tiempo t. Permite esto darle interpretación a μ como una media global.
- Como es clásico se asume que los errores tienen distribución normal y son independientes

$$\varepsilon_{i,1,z} \overset{\text{i.i.d}}{\sim} N\left(0, \frac{\sigma_\varepsilon^2}{1-\phi^2}\right), \quad \varepsilon_{i,j,z} \overset{\text{i.i.d}}{\sim} N\left(0, \frac{\sigma_\varepsilon^2\left(1-\phi^{2\gamma(i,j,z)}\right)}{1-\phi^2}\right)$$

El modelo subyacente es $u_{i,j,z} = y_{i,j-1,z} - \mu - \beta_{t(i,j-1,z)} - \tau_z$, entonces podemos reescribir la serie como,

$$u_{i,j,z} = \phi^{\gamma(i,j,z)} u_{i,j-1,z} + \varepsilon_{i,j,z}$$

Estamos en presencia de un modelo autorregresivo estacionario donde la covarianza entre dos ventas del mismo inmueble es,

$$cov\left(u_{i,j,z}, u_{i,j-1,z}\right) = \sigma_\varepsilon^2 \frac{\phi^{(t(i,j',z)-t(i,j,z))}}{1-\phi^2} \quad \text{si } j < j'.$$

Por tanto la dependencia entre las ventas de un mismo inmueble va disminuyendo a medida que crece el tiempo entre las tasaciones. También cabe destacar que a medida que el tiempo entre una venta y su anterior aumenta, la variabildad del error también crece, lo cual indica que cuanto más tiempo a pasado de la venta anterior el valor de dicha venta posee menos información acerca del valor actual del inmueble.

2.6 Método de Stacking

El método de regresión de Stacking introducido por Wolpert, [Wolpert, 1992], posteriormente desarrollado por Leo Breiman, [Breiman, 1996], consiste en la construcción de un nuevo predictor a partir de una combinación lineal de K predictores conocidos.

Supongamos que conocemos K predictores $v_1(\mathbf{x}),\ldots,v_K(\mathbf{x})$ de un ouput $y \in \mathbb{R}$, donde $\mathbf{x} \in \mathbb{R}^d$. Si llamamos $\mathscr{L} = \{(y_i,\mathbf{x}_i)/i = 1,\ldots,N\}$ al conjunto de las muestras de entrenamiento, donde $\mathbf{x}_i$ es el vector de entrada, en nuestro caso puede suceder que dicho vector tenga datos faltantes en algunas de sus componentes para algún i, es decir, se presenta un problema de regresión con datos faltantes.

Combinemos estos predictores de forma lineal para la construcción de un nuevo predictor v,

$$v(x) = \sum_k \alpha_i v_i(\mathbf{x}) \tag{12}$$

Una posible estimación de los coeficientes $\{\alpha_i\}$ es usar mínmos cuadrados ordinarios, es decir aquellos coeficiente que minimizan el error en L_2. Esta metodología puede ocasionar un sobreajuste a los datos y además en general existe el problema de la alta correlación de los predictores, por tanto estimaciones mediante penalizaciones, por ejemplo regresión Lasso, suelen ser mas adecuadas. Una forma de garantizar que la predicción se encuentre en el rango $[\text{mín}_k\, v_k(\mathbf{x}), \text{máx}_k\, v_k(\mathbf{x})]$, es imponer restricciones de convexidad, es decir,

$$\sum_k \alpha_k = 1 \quad \text{y} \quad \alpha_k \geq 0. \tag{13}$$

3 Descripción de la base de datos

La base de datos fue proporcionada por el Banco Hipotecario del Uruguay (BHU). La variable de interés es el valor contado de un inmueble en el Departamento de Montevideo, Uruguay. Dada una base donde contamos con el valor contado de determinados inmuebles en diferentes momentos de tiempo, pudiendo o no tener el valor contado actual (entendemos por valor contado actual aquel que fue fijado posterior al 1 de octubre de 2014), su ubicación espacial, ciertas variables intrínsecas al inmueble (como por ejemplo la superficie, el número de dormitorios, antigüedad, estado de la propiedad) y para algunos de dichos inmuebles se conoce su valor contado en uno o algunos momentos temporales. El valor contado del inmueble es dado por un tasador dependiente del BHU. En este trabajo no se considera el efecto tasador. Como forma de simplificar el modelo las variables hedónicas del bien se consideran no modificables a través del tiempo. Por tanto aquellos inmuebles que no cumplan este requerimiento, que presenten datos faltantes en la covariables del modelo, o alguna inconsistencia respecto a la fecha de tasación o al valor de la tasación contado son descartadas de la base. Si llamamos “tasaciones hermanas” a aquellas inmuebles tasados en la misma fecha, en el mismo padrón, con igual valor contado. Consideramos que corresponden a apartamentos gemelos de los cuales sólo uno fue tasado. En este caso nos quedamos con uno sólo de los inmuebles. El modelo tiene por objetivo, dado un inmueble, en cierta ubicación, con determinadas características propias de la propiedad, pudiendo o no ser conocidos valores contados de años anteriores, asignar un precio a la propiedad. Es claro que esta metodología puede proporcionar un precio referencial para el inmueble así como un intervalo de confianza para dicho valor, pero este debe ser asignado y confirmado por el tasador. Cabe aclarar que en el trabajo trataremos como sinónimos el valor contado con el valor de tasación. Se modelan las tasaciones entre 5,000 y 500,000 de dólares. Se consideran aquellos barrios en los que el BHU tiene al menos 20 inmuebles. Luego de depurada la base se cuenta con 20422 registros de 12732 inmuebles, desde el 8 de marzo de 1973 al 23 de junio de 2015. En la tabla 3 se observa que 7763 tienen una única tasación.

Número de Tasaciones	Inmuebles
1	7763
2	2916
3	1562
4	363
5	91
6	29
7	4
8	4

En el mapa de la figura 1 se encuentra la ubicación de los inmuebles con el valor de su última tasación registrada en cualquier fecha. Consideramos como tasación vigente aquella que fue realizada luego del 1 de octubre del 2014. En el mapa de la figura 2 se muestra la ubicación de los 1621 inmuebles con tasación vigente. En el apéndice 5 se registra, de las tasaciones vigentes, el valor contado promedio por barrio.

En los mapas de la figura 3 se representa la ubicación de aquellas propiedades donde su última tasación fue superior a 300,000 dólares o inferior a 20,000 dólares respectivamente.

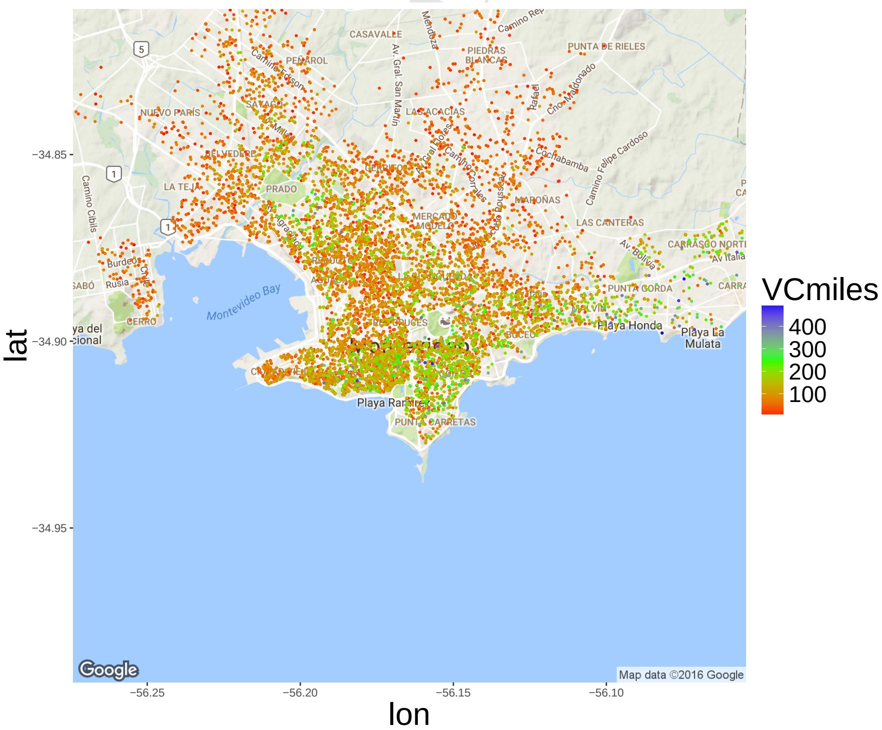

Figura 1: Ubicación y valor contado (última tasación registrada) de los 12732 inmuebles.

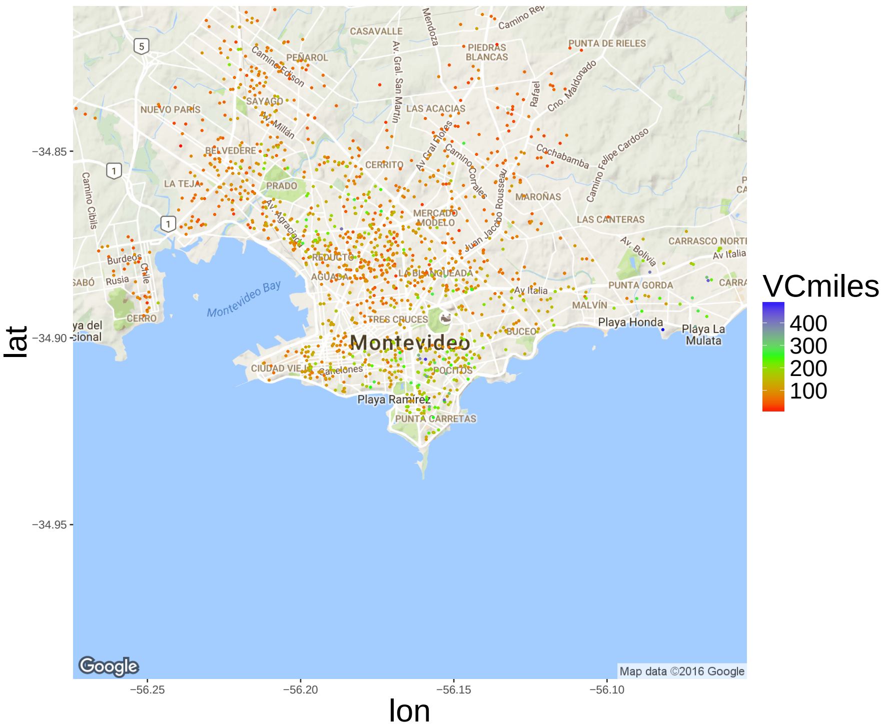

Figura 2: Ubicación y valor contado de los 1621 inmuebles con tasación vigente.

Es claro como el precio de los inmuebles decrece desde la costa hacia el norte. Sin embargo barrios como el Prado presentan un comportamiento particular. Los inmuebles de mayor valor en general se ubican en los barrios de Carrasco y Punta Gorda, mientras que los de menor valor se ubican en la periferia.

Es clara la dependencia temporal y como influyeron las crisis de principios de 1980 y principios del 2000 en el precio promedio de las propiedades, ver figura 4. Si observamos el lapso de tiempo en años entre tasaciones de una misma propiedad vemos que en general varia entre 1 y 5 años, pero puede pasar inclusive más de 20 años entre una tasación y la siguiente, ver figura 4.

En la figura 5, se describen las covariables número de dormitorios, antigüedad de la propiedad, superficie y fecha de tasación.

4 Resultados

Se consideran 200 inmuebles elegidos al azar como muestra de testeo. Como resultado inicial y de forma de poder desarrollar todos los modelos, se toman en la muestra de testeo inmuebles que tengan al menos una tasación no vigente.

Luego de implementados los modelos se calcula como una medida de performance la raíz del error cuadrático medio (anotamos *RECM*), es decir, si Y_i es el valor contado en la muestra de testeo y $\hat{Y}_i$ su estimación,

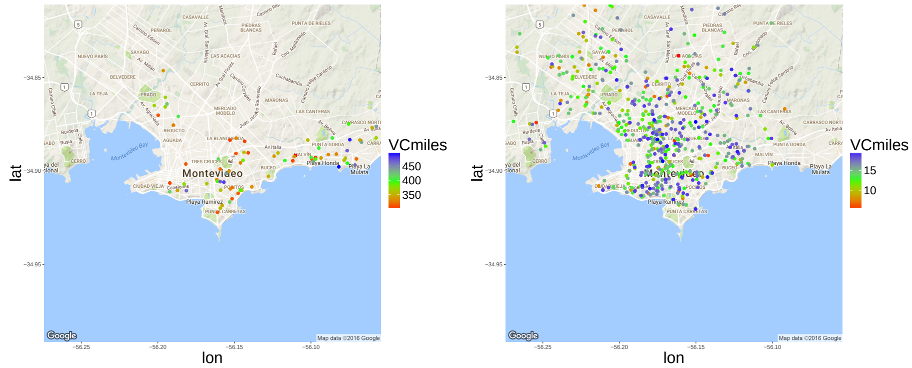

Figura 3: Ubicación y valor contado (última tasación registrada) de aquellos inmuebles cuyo valor contado esta por encima de 300.000 dólares y por debajo de 20,000 dólares respectivamente.

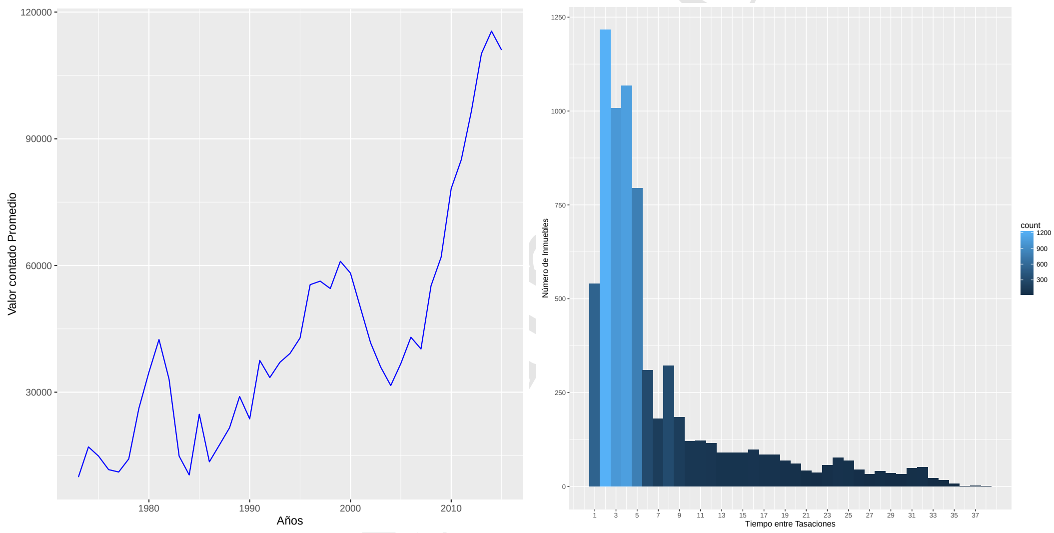

Figura 4: Valor promedio contado de todas las tasaciones por año y tiempo en años entre una tasación y la siguiente.

$$RECM = \sqrt{\frac{1}{n}\sum_{i=1}^{300}\left(Y_i - \hat{Y}_i\right)^2} \tag{14}$$

En Stacking se consideraron cinco regresores,

- El Modelo Semiparamétrico Espacial Dinámico.
- El Modelo de Bailey, Muth y Nourse.
- El Modelo de Case y Shiller.
- Modelo de Efectos Mixtos.
- El Modelo No-paramétrico del Indice Temporal.

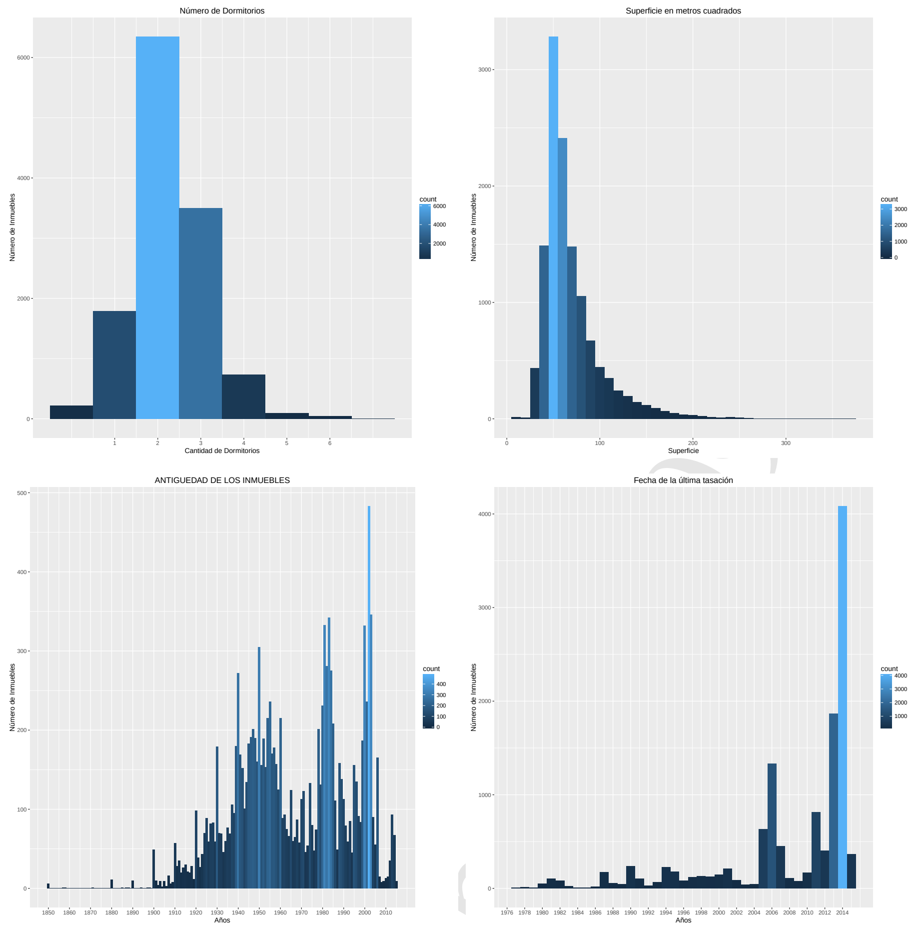

Figura 5: Gráficos descrpitvos de algunas de las covariables y de la fecha de tasación de los inmuebles.

Los resultados obtenidos se muestran en la tabla 1.

Es posible en los distintos modelos observar las estimación del índice de precios y la diferencia entre la predicción y el valor real en la muestra de testeo, por ejemplo en el modelo no paramétrico, ver figura 6, se observa claramente la crisis de principios del 2000 pero no así la principios de 1980.

5 Conclusiones

Como se aprecia en los resultados el RECM disminuye al aplicar la metodología de Stacking produciendo predicciones más precisas. Es evidente que todos los modelos que se basan en ventas repetidas son más eficientes, sin embargo estos modelos no pueden ser utilizados cuando no tenemos ninguna tasación previa del inmueble. Como trabajo futuro se podrían considerar otras muestras de testeo y evaluar el error cuadrático medio promedio. Por otro lado, un objetivo más ambicioso, es intentar construir un modelo que tenga como insumo toda la información simultáneamente y no usar técnicas agregativas. Creemos que la combinación de lo realizado en [Nagaraja et al., 2011] y [Sun et al., 2014] sería el camino apropiado para la construcción de este modelo.

Metodología	RECM
Modelo Semiparamétrico Espacial Dinámico	24789
Modelo de Bailey, Muth y Nourse	18987
Modelo de Case y Shiller	15622
Modelo de Efectos Mixtos	21897
Modelo Autoregresivo de Nagaraja	13234
Modelo No-paramétrico del Indice	14355
Stacking	12232

Tabla 1: Raíz del error cuádratico medio en la muestra de testeo a partir de los distintos modelos.

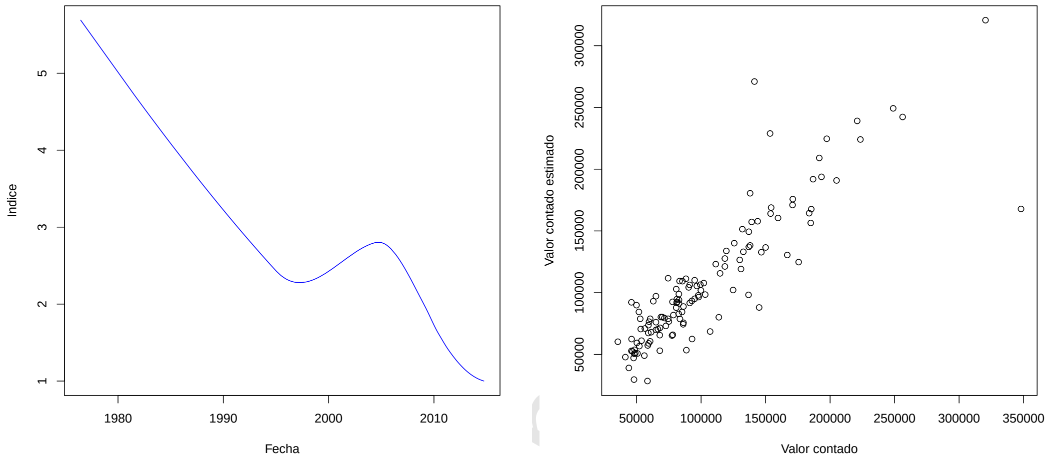

Figura 6: Estimación de la función índice de precios y gráfico de los valores contados en referencias a sus predicciones en la muestra de entrenamiento.

Referencias

[Anselin, 1988] Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, Dordrecht.

[Bailey et al., 1963] Bailey, M., Muth, R., and Nourse, H. (1963). Regression method for real estate price index construction. *Journal of the American Statistical Association*, 58:933–942.

[Breiman, 1996] Breiman, L. (1996). Stacked Regressions. *Machine Learning*, 24:49–64.

[McCulloch and Searle, 2001] McCulloch, C. E. and Searle, S. R. (2001). *Generalized, Linear, and Mixed Models*. John Wiley and Sons, New York.

[Nagaraja et al., 2010] Nagaraja, C. H., Brown, L., and Wachter, S. (2010). House Price Index Methodology. Technical report.

[Nagaraja et al., 2011] Nagaraja, C. H., Brown, L. D., and Zhao, L. H. (2011). An autoregressive approach to house price modeling. *The Annals of Applied Statistics*, 5(1):124–149.

[Shiller and Case, 1987] Shiller, R. J. and Case, K. E. (1987). Prices of Single Family Homes Since 1970: New Indexes for Four Cities. *New England Economic Review*, Sept/Oct:45–56.

[Sun et al., 2014] Sun, Y., Yan, H., Zhang, W., and Lu, Z. (2014). A semiparametric spatial dynamic model. *The Annals of Statistics*, 42(2):700–727.

[Wolpert, 1992] Wolpert, D. H. (1992). Stacked Generalization. *Neural Networks*, 5:241–259.

Apéndice 1: Valor promedio de los inmuebles por barrio.

BARRIO	Media
LAVALLEJA	41888.20
CASAVALLE	43617.00
JARDIN DEL HIPODROMO	49554.00
NUEVO PARIS	56013.38
HIPODROMO	60400.33
BELLA ITALIA	60437.25
FLOR DE MAROÑAS	60490.47
VICTORIA	60546.33
PIEDRAS BLANCAS	61161.15
LEZICA	62937.62
PASO DE LA ARENA	63200.88
ITUZAINGO	63946.23
CORDON NORTE	68702.00
CERRITO DE LA VICTOR	70027.00
CERRO	70195.17
CONCILIACION	71696.20
PEÑAROL	71944.87
CIUDAD VIEJA	72866.38
VILLA ESPA NOLA	73149.00
MANGA	74284.80
LA TEJA	75368.68
MAROÑAS	75467.85
BELVEDERE	76483.82
PASO MOLINO	76572.12
RETIRO	78318.73
PARQUE DEL SOL	79455.00
CAPURRO	81092.39
MALVIN NORTE	81193.45
VILLA MUÑOZ	81671.33
FERROCARRIL	83275.75
UNION	84978.40

BARRIO	Media
SAYAGO	86451.01
ABAYUBA	89229.83
PEREZ CASTELLANO	90290.46
LA COMERCIAL	91779.56
REDUCTO	94913.29
COLON	95793.76
JACINTO VERA	98474.33
TRES CRUCES	99750.00
ARROYO SECO	103600.00
CORDON SUR	103849.67
AIRES PUROS	103923.18
LARRAÑAGA	104151.54
BOLIVAR	105114.33
SUR	106583.33
GOES	106905.29
CENTRO	107623.92
LA BLANQUEADA	108788.25
AGUADA	111233.67
LA FIGURITA	114532.56
BUCEO	116523.70
PARQUE BATLLE	123272.09
BELLA VISTA	124989.71
VILLA DOLORES	125800.00
CORDON	126851.93
PARQUE POSADAS	128116.00
BRAZO ORIENTAL	129581.68
ATAHUALPA	131734.60
PRADO	133205.62
MALVIN	145217.54
PALERMO	149623.25
PARQUE RODO	153275.30
MALVIN NUEVO	156421.75
PUNTA CARRETAS	173061.40
POCITOS	174356.54
CARRASCO NORTE	196198.60
CARRASCO	229272.27
PUNTA GORDA	253036.36