

> Año 10 / Número 101 / Marzo 2011

Economía de las ciudades

Efecto de la cercanía a las estaciones de Subte y valor de la propiedad residencial en Buenos Aires

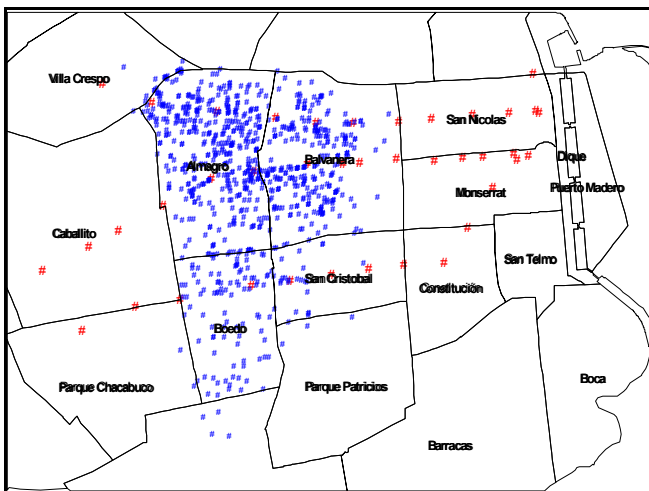
Un estudio según el modelo de valuación hedónico | Por Marcelo Oscar Castillo

F) Metodología

a) Tratamiento dado a los datos

En primer lugar, se seleccionaron los avisos utilizables, los que debían contener información al menos sobre la dirección exacta de la propiedad, precio y superficie en m². Para los avisos “utilizables” no solamente se cargaron los datos mencionados sino también todos los datos incluidos en el aviso.

Luego se georreferenciaron los avisos mencionados en el párrafo anterior y se generó una variable llamada “distancia a estación del subte”.



Se realizaron regresiones de corte transversal, buscando evaluar la significatividad de los coeficientes para diferentes atributos de los departamentos, siendo “distancia a estación del subte” uno de ellos.

b) Determinación de la forma funcional

Como bien notan Meloni y Ruiz Nuñez (1998), la teoría económica disponible no especifica la forma funcional que debe adoptarse para explicar el valor de las propiedades. Es decir, partimos de la suposición de que el precio de las propiedades guarda relación con una serie de atributos. Así, tenemos que: $P=f(\text{Distancia a las estaciones del subte, superficies, antigüedad, barrio, Patio(SI/NO), Terraza(SI/NO), Balcón (SI/NO), cochera (SI/NO), PH (SI/NO), Por Escalera (SI/NO)}.$

(Donde P es el precio de la propiedad o una función de dicha variable). Sin embargo, es necesario elegir cual o cuales van a ser las formas funcionales de la relación arriba citada para poder realizar las regresiones y cuantificar la influencia de cada una de las variables sobre el precio de las propiedades.

En este trabajo, para definir la forma funcional se consideró un modelo lineal de la forma $yY = \alpha + \beta g(x) + \varepsilon$ y se realizó mediante el soft Stata la transformación de Box Cox, cuya expresión es

$$g^\lambda(x) = \begin{cases} (x^\lambda - 1) / \lambda & \text{si } \lambda \neq 0 \\ \text{Ln}(x) & \text{si } \lambda = 0 \end{cases}$$

La transformación de Box Cox permite que los datos “sugieran” la forma funcional más apropiada, sin tener que definir a priori el tipo de relación funcional que se va a usar.

El modelo lineal se da cuando $\lambda=1$, mientras que los modelos log-lineal o semilogarítmicos (dependiendo de cómo se haya medido Y) se dan si $\lambda=0$. Otros valores de λ dan lugar a otras muchas formas funcionales diferentes. Por ejemplo, si $\lambda=-1$ entonces la ecuación incluirá la recíproca de X (variable explicativa). El modelo Box Cox es una transformación útil que permite una gran flexibilidad al incluir a λ como un parámetro adicional desconocido. El coste de esta flexibilidad es que el modelo se transforma en no lineal en los parámetros.

En principio, cada regresor podría transformarse con un valor de λ diferente pero en la mayoría de las aplicaciones este nivel de generalidad se transforma en algo incómodo y se suele usar un mismo valor de λ para todas las variables explicativas. Simultáneamente se puede transformar y utilizando otro parámetro.

Además, en la mayoría de las aplicaciones las variables dummy no son transformadas.

El resultado de la transformación Box Cox fue el siguiente:

Number of obs = 820
 LR chi2(13) = 1584.68
 Log likelihood = -8394.0266 Prob > chi2 = 0.000

precio	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
/lambda	-.0385	.0533469	-0.72	0.470	-.1431217	.0659944
/theta	-.176	.0500098	-3.54	0.000	-.2749345	-.0788998

Test	Restricted	chi2	Prob > chi2
H0:	log likelihood		
theta=lambda = -1	-8517.5241	247.00	0.000
theta=lambda = 0	-8401.1406	14.23	0.000
theta=lambda = 1	-8687.4727	586.89	0.000

Y usando igual parámetro para ambos lados del modelo de regresión los resultados fueron:

LR chi2(12) = 1577.53
 Log likelihood = -8397.6024 Prob > chi2 = 0.000

precio	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
/lambda	-.1218424	.0469794	-2.59	0.009	-.2139204	-.0297644

Test	Restricted	LR statistic	P-Value
H0:	log likelihood	chi2	Prob > chi2
lambda = -1	-8517.5241	239.84	0.000
lambda = 0	-8401.1406	7.08	0.008
lambda = 1	-8687.4727	579.74	0.000

Se observa que si bien todas las hipótesis se rechazan la hipótesis que está más cerca de ser aceptada es la de $\lambda = 0$ (es decir, la correspondiente a un modelo log – log)

En función de los resultados obtenidos se decidió usar una forma funcional del tipo log-log.

Un aspecto positivo de la forma funcional elegida es que las interpretaciones de los coeficientes son más intuitivas, ya que se tratan de elasticidades o semielasticidades.

Esto es beneficioso sobre todo si se tiene en cuenta la heterogeneidad de las propiedades: si los coeficientes reflejaran valores absolutos no resulta claro si un aumento del precio de, por ejemplo, 3.000 dólares para los departamentos con cochera es grande o pequeño. En cambio resulta mucho más ilustrativo un coeficiente que indica que los departamentos con cochera se venden un 5%

más caros. Como antecedente pueden citarse los trabajos de Conte Grand (2001) y Meloni y Ruiz Nuñez (1998), donde también se utilizaron modelos log log.

c) Tratamiento de los sesgos de selección

Se analizaron los posibles sesgos de selección de la muestra, concluyéndose que no existen razones para esperar sesgo de selección por la fuente de información elegida pero que es posible que la metodología seguida introduzca un sesgo en contra de los departamentos a estrenar.

El requisito de que los avisos incluyan la superficie y el precio para ser incluidos en la base de datos podría en parte limitar la aparición de edificios nuevos en la muestra, ya que la superficie no suele ser informada en los avisos de departamentos nuevos.

El tratamiento correcto de tal sesgo de selección implicaría la utilización de procedimientos como los sugeridos por Heckman (1979). En su trabajo "Sample selection as a specification error" James Heckman presenta un procedimiento de estimación dividido en dos etapas; en la primera, se estima una ecuación de selección incorporando variables que afecten la probabilidad de observar la variable dependiente relevante del modelo. De esta estimación, se obtiene un término de corrección que se incorpora en la segunda etapa o ecuación de regresión, conjuntamente con el resto de las variables de interés. El tratamiento de posibles sesgos de selección excede el alcance de este trabajo, fundamentalmente por el tiempo que demandaría. En efecto, sería necesario incluir en la base todas las propiedades publicadas y utilizarlas en la regresión de la ecuación de selección.

El no tratamiento de los eventuales sesgos de selección implica que las conclusiones del estudio son válidas únicamente para la muestra seleccionada y no para la población.

d) Tratamiento a dar a la inflación

Dado que se trabajó con un período de tiempo relativamente breve (3 meses) no se tomaron medidas para contemplar la inflación del precio de las propiedades. De todas formas se revisaron los datos, observándose que las viviendas que aumentaron su valor en el período fueron escasas, prevaleciendo en forma general el descenso del valor de las propiedades presumiblemente al no encontrar compradores luego de las primeras publicaciones.