

> 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

G) Resultados

Los modelos que se presentan en esta sección no incluyen muchas de las variables enumeradas en el [Anexo I](#). Esto se debe a que las mismas resultaron no significativas en todas o casi todas las regresiones realizadas. Las variables patio y terraza resultaron significativas pero fueron eliminadas porque tenían signos distintos a los esperados y porque se comprobó que la estimación de los coeficientes de las otras variables no cambiaba al eliminarlas. El R² prácticamente no varió.

Heterocedasticidad

Se realizaron los test de White y de Goldfeld Quant para detectar la presencia de heterocedasticidad de los residuos. En ambos casos se rechazó la hipótesis nula de homocedasticidad.

La heterocedasticidad (heterogeneidad de la varianza del término de error) tiene implicaciones potencialmente serias para inferencias basadas en los resultados de mínimos cuadrados ordinarios.

La heterocedasticidad no destruye el insesgamiento ni las propiedades de consistencia de los estimadores de MCO. Sin embargo, en presencia de heterocedasticidad no poseen varianza mínima y por lo tanto no son eficientes.

Si se continúa utilizando los estimadores usuales de MCO en casos de heterocedasticidad las varianzas de estos estimadores dejan de ser calculadas por medio de las fórmulas habituales de MCO.

Si se continúan utilizando los estimadores de MCO para las varianzas de los estimadores de los coeficientes las estimaciones, además de ineficientes, dejan de ser insesgadas y las pruebas t y F llevarán a conclusiones erradas.

En el trabajo la forma de tomar en cuenta la heterocedasticidad fue utilizar el estimador de White (1980), que permite estimar la varianza de los estimadores aún sin conocer la forma de la heterocedasticidad. Ha habido numerosos estudios que buscan mejorar el estimador de White para mínimos cuadrados ordinarios. Las propiedades asintóticas del estimador son ambiguas y su utilidad en pequeñas muestras suscita muchas incógnitas. La raíz de los posibles problemas parten del resultado general de que los cuadrados de los residuos MCO tienden a subestimar los cuadrados de las verdaderas perturbaciones. El problema de esto es que el estimador de White sería demasiado optimista en muestras pequeñas y entonces los valores t son demasiado altos.

Existen otras técnicas, cuyo uso permitiría una mejor estimación de los parámetros del modelo. Davidson y MacKinnon (1993) sugieren una serie de soluciones que incluyen: 1) multiplicar el resultado final por un factor $n/(n-K)$ (donde n son las observaciones y K los parámetros) y 2) usar los cuadrados de los residuos transformados por su verdadera varianza, $e_i^2 m_{ij}$ en vez de e_i^2 , donde $m_{ij} = 1 - x_i' (X'X)^{-1} x_i$. Según su estudio, Davidson y MacKinnon son partidarios de una u otra corrección.

A continuación se presentan los resultados de las estimaciones utilizando el estimador de White para corregir la heterocedasticidad (en el [Anexo II](#) se incluyen los resultados de las regresiones sin utilizar el estimador de White).

1) Modelo considerando la influencia de la distancia como anillos a 400 y 800 metros

logprecio	Coef.	Robust Std. Err.	t	P> t	[95% Conf.	Interval]
-----	-----	-----	-----	-----	-----	-----
cochera	0,107	0,019	5,42	0	0,068	0,147
A_reciclar	-0,127	0,038	-3,29	0,001	-0,203	-0,051
por_escale	-0,051	0,021	-2,34	0,019	-0,093	-0,008
Almagro	0,105	0,011	9,42	0	0,083	0,127
Ph	-0,112	0,019	-5,77	0	-0,150	-0,074
subte0400	0,144	0,039	3,62	0	0,065	0,222
Subte400800	0,130	0,040	3,2	0,001	0,050	0,207
misereresino	-0,122	0,041	-2,92	0,004	-0,204	-0,040
Balcón	0,077	0,011	6,67	0	0,054	0,099
Supdescsino	0,142	0,040	3,54	0	0,063	0,222
Log(Antigüedad)	-0,050	0,006	-7,41	0	-0,064	-0,037
Log(Superficie)	0,765	0,015	50,72	0	0,736	0,795
cons	7,454	0,075	98,99	0	7,306	7,602

F(12,807)= 340,02
R2= 0,8540

2) Modelo considerando la influencia de la distancia como anillos a 500 y 1000 metros

Logprecio	Coef.	Std. Err.	t	P>t	[95% Conf. Interval]	
Cochera	0,107	0,020	5,37	0.000	0,068	0,146
A_reciclar	-0,123	0,038	-3,24	0.001	-0,198	-0,048
por_escale	-0,048	0,021	-2,24	0.025	-0,090	-0,005
Almagro	0,105	0,011	9,39	0.000	0,083	0,127
Ph	-0,116	0,019	-5,88	0.000	-0,154	-0,077
misereresino	-0,122	0,041	-2,93	0.003	-0,204	-0,040
balcontotal	0,077	0,011	6,71	0.000	0,054	0,100
Supdescsino	0,145	0,039	3,63	0.000	0,066	0,223
Log (Antigüedad)	-0,05	0,006	-7,33	0.000	-0,064	-0,037
Log (Superficie)	0,763	0,015	50,28	0.000	0,733	0,793
subte0500	0,127	0,042	2,99	0.003	0,043	0,210
subte5001000	0,097	0,043	2,22	0.027	0,011	0,183
cons	7,479	0,077	97,01	0.000	7,328	7,630

F(12, 807) = 351,17
R2 = 0,8534

3) Modelo considerando la influencia de la distancia como una variable continúa (se tomaron logaritmos de la distancia a las estaciones)

Logprecio	Coef.	Std. Err.	t	P> t	[95% Conf.	Interval]
Cochera	0,108	0,020	5,37	0	0,068	0,148
A_reciclar	-0,122	0,038	-3,19	0,001	-0,197	-0,047
por_escale	-0,046	0,021	-2,17	0,03	-0,088	-0,004
Almagro	0,112	0,011	10,2	0	0,091	0,134
Ph	-0,117	0,019	-6,08	0	-0,155	-0,079
Log(Distancia a Estacion)	-0,026	0,007	-3,32	0,001	-0,042	-0,010
misereresino	-0,116	0,041	-2,81	0,005	-0,198	-0,035
Balcon	0,076	0,011	6,62	0	0,053	0,099
supdescsino	0,152	0,039	3,85	0	0,074	0,229
Log(Antigüedad)	-0,052	0,006	-7,48	0	-0,065	-0,038
Log(Superficie)	0,761	0,014	50,79	0	0,732	0,790
cons	7,759	0,077	99,8	0	7,606	7,912

F(11, 808) = 380,39
R2 = 0,852

Se observa que las variables relacionadas con la distancia a las estaciones continúan siendo claramente significativas. Sin embargo, los resultados

muestran un hecho que podría estar reflejando un problema de la forma funcional elegida: mientras que el modelo 1 muestra que pertenecer al anillo entre 0-400 metros o al de 400-800 metros no hace una gran diferencia en lo que se refiere al precio de las propiedades (14,4% y 13%, respectivamente, por encima de los precios de las propiedades situadas a más de 800 metros de distancia de las estaciones), el modelo 3 muestra una elasticidad del precio respecto a la distancia a las estaciones de $-0,026$ (es decir, que ante un aumento de 1% en la distancia a las estaciones el precio de la propiedad caería 2,6%). Es decir, el modelo 1 estaría mostrando una influencia pequeña de la distancia a las estaciones, en tanto que el modelo 3 mostraría que tal influencia es muy significativa.

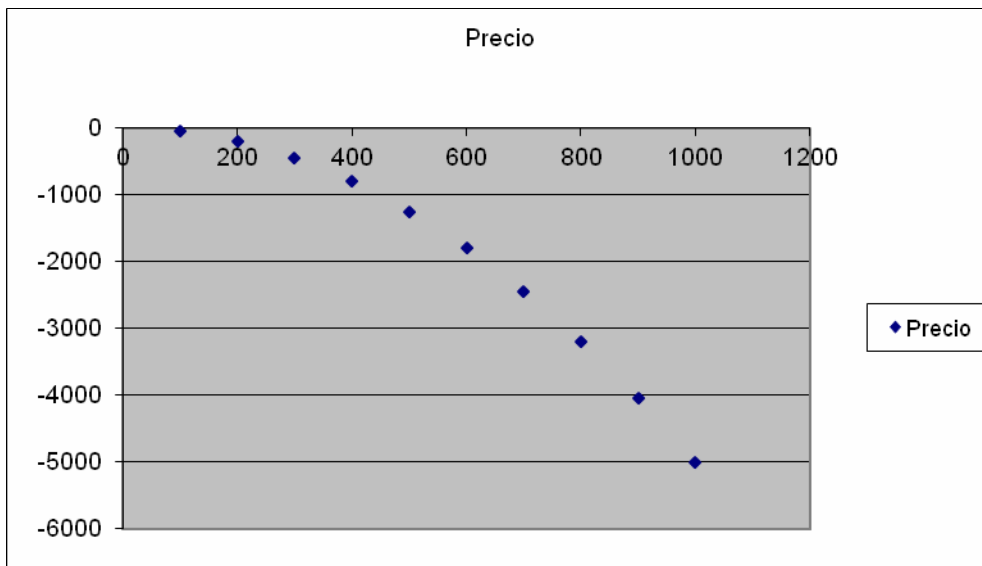
Ante esta situación se decidió probar con una forma funcional cuadrática, que sería la apropiada si los resultados del modelo 1 fueran los más cercanos a la realidad.

4) Modelo lineal tomando el cuadrado de la distancia a las estaciones de subte

		Robust				
Precio	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
Cochera	8.069	1.971	4,09	0.000	4.199	11940
a_reciclar	-4.700	2.020	-2,33	0.020	-8.666	-7.347
por_escale	-1.194	1066	-1,12	0.263	-3.287	8.994
Antigüedad	-116	26	-4.40	0.000	-1.678	-6.426
Almagro	5.188	680	7,63	0.000	3.853	6.523
Ph	-4.824	1.243	-3,88	0.000	-7.265	-2.382
Misereresino	-7.059	2.406	-2,93	0.003	-11783	-2.335
Balcontotal	4.282	749	5,71	0.000	2.811	5754
Supdescsino	8.627	2.780	3,10	0.002	3.169	14085
Supregr	479	22	21,38	0.000	435	523
Cuadrado de la distancia a estaciones	-0,005	0,001	-3,16	0.002	-,008	-,001
_cons	14105	1.478	9,54	0.000	11203	17007

$F(11,808) = 97,44$

$R^2 = 0,7897$



Se observa que la suave caída del precio en los primeros 400 – 500 metros parece confirmar el análisis realizado tomando a la distancia en forma de anillos.

Multicolinealidad

Existe multicolinealidad cuando se verifica que las variables explicativas están intercorrelacionadas, de la siguiente forma:

$$\lambda X_1 + \lambda X_2 + \lambda X_3 + \lambda X_n + v = 0$$

(donde λ son todas constantes tal que no todas sean simultáneamente 0 y v es un término estocástico de error)

Si bien es cierto que los estimadores MCO continúan siendo insesgados si existe colinealidad debe recordarse que el insesgamiento es una propiedad de muestras repetidas. Es decir, no podemos asegurar que el insesgamiento se mantenga para una muestra dada.

También es cierto que la multicolinealidad no destruye la propiedad de varianza mínima de los estimadores MCO, pero eso no significa que tales varianzas sean pequeñas para una muestra dada.

La multicolinealidad es un fenómeno muestral. Por lo tanto, puede ocurrir que variables que no estén correlacionadas en la población lo estén en una muestra dada. Incluso puede ocurrir que la correlación sea tan grande que impida aislar la influencia de las variables sobre Y .

Por todo esto, el hecho de que los estimadores MCO sean los mejores estimadores lineales insesgados a pesar de la multicolinealidad es poco importante en la práctica. Esto implica que debemos ver qué es lo que puede ocurrir en una muestra dada.

En una muestra dada las consecuencias pueden ser las siguientes:

- Varianzas y covarianzas amplias para los estimadores MCO
- Intervalos de confianza más amplios
- Razones t no significativas
- Un valor elevado de R2 pero pocas razones t significativas
- Sensibilidad de los estimadores MCO y de sus desvíos estándar ante cambios pequeños en los datos.

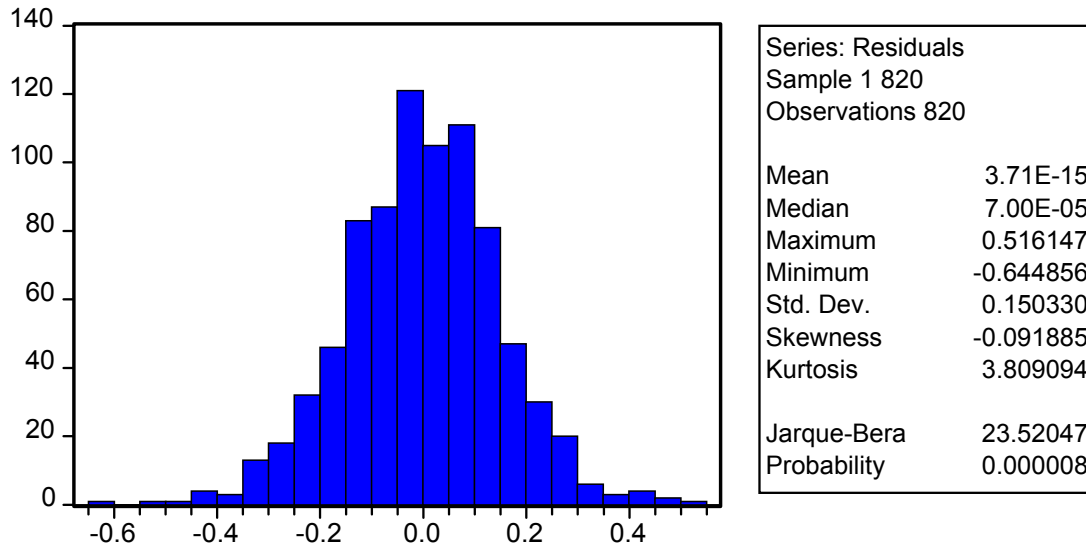
En el trabajo se comprobó que las variables “cantidad de ambientes” y logsupregr estaban altamente correlacionadas y entonces se eliminó la variable “cantidad de ambientes” del modelo.

Se comprobó que cuando se incluía la variable “Cantidad de Ambientes” en el modelo el error estándar de la variable logsupregr es el doble que cuando no se utiliza la variable “ambientes” y además el valor del coeficiente es muy diferente (0,59 vs. 0,76). La correlación entre ambientes y logsupregr es 0,8306.

No se detectaron otras variables correlacionadas.

Normalidad de los residuos

Se efectuó el test de Jarque Bera para chequear el cumplimiento del supuesto de normalidad de los residuos (para el modelo 1). A continuación se incluye la salida del soft Eviews 3.



El test de Jarque Bera indica que se rechaza la hipótesis de normalidad de los residuos. Se observa que si bien la media es prácticamente igual a cero la kurtosis difiere bastante del valor correspondiente a una distribución normal (3,80 contra 3).

Siguiendo a Gujarati (1997) tenemos que si los residuos no están distribuidos normalmente distribuidos una extensión del teorema central del límite permite aceptar que los coeficientes estarán normalmente distribuidos en forma

asintótica y entonces podrán efectuarse las pruebas usuales (t y F). Para que esto ocurra deben darse las siguientes condiciones:

- Que los residuos sean independientes e idénticamente distribuidos con media igual a cero y varianza constante.
- Que las variables explicativas sean constantes en muestras repetidas
- Que las muestras sean grandes.

Lamentablemente se ha visto que los residuos son homocedásticos, lo que hace que las estimaciones de los coeficientes por MCO no puedan considerarse confiables.

Sin embargo debe observarse que la distribución es más leptocúrtica que la normal, por lo que la situación no sería tan mala: cualquier hipótesis nula no rechazada en una distribución normal tampoco lo sería en la distribución real (leptocúrtica) de los residuos. Además, debe tenerse en cuenta que los valores t obtenidos son altamente significativos.

Outliers

El método utilizado para detectar outliers fue el análisis de los residuos estandarizados. Se siguió la regla empírica según la cual se consideran outliers a aquellos residuos estandarizados cuyo valor z es mayor que 3,5 o menor que -3,5.

Se encontraron outliers en dos observaciones, con valores de -3,63 y -4,288 para los residuos estandarizados. En ambos casos se trataba de propiedades de gran antigüedad y superficies grandes, con posibilidades de que se incluyeran grandes superficies descubiertas en el valor de la variable superficie.

Se probó el modelo 1 eliminando las dos observaciones sin obtener cambios significativos en los coeficientes y en el R² (los resultados del modelo eliminando los outliers se incluyen en el [Anexo II](#)).