

FUNCIÓN DE PRECIOS HEDÓNICOS EN EL MERCADO DE LA VIVIENDA

Iván

ivancorujo@h...

Resumen

El precio de la vivienda afecta tanto a los individuos que demandan servicios de vivienda como a los que desean comprar una. La pregunta que se desea responder en este trabajo es, ¿Qué factores determinan el precio de la vivienda y en qué magnitud? Existe un enfoque teórico denominado **Modelo de precios hedónicos**, que afirma que el precio de la vivienda está en función directa de sus características, como por ejemplo: la cantidad de habitaciones, calidad de la construcción, metros cuadrados, etc. Estas características del mercado de vivienda son esenciales porque llevan a un tratamiento diferente con respecto del que se lleva a cabo en otros mercados. El objetivo principal del trabajo es estimar, a través de herramientas econométricas, la función de precios hedónicos sobre la base de características asociadas a la construcción y localización de las viviendas.

Palabras claves: mercado inmobiliario, determinación de precios, precios hedónicos, econometría, regresión, economía.

Abstract

The price of housing affects both individuals who demand housing services, and those who want to buy one. The question to be answered in this work is, what factors determine the price of housing and in what magnitude? There is a theoretical approach called hedonic pricing model, which states that the price of housing is a direct function of their characteristics, such as: number of rooms, quality of construction, square footage, etc.. These features of the housing market are essential because they lead to a different treatment in respect of which is carried out in other markets. The main objective of the study is to estimate, through econometric tools, a hedonic pricing function on the basis of characteristics associated with the construction and location of housing.

Keywords: real estate market, pricing, hedonic pricing, econometrics, regression, economy.

Introducción

Como introducción se describen algunas características de las viviendas, las cuales sirven para dar una idea del mercado inmobiliario.

-Heterogeneidad: las viviendas se distinguen entre ellas en muchos aspectos lo están asociados a diferentes características en su construcción (cantidad de ambientes, tamaño de la cocina y del parque, metros cuadrados, antigüedad, etc.).

-Inmovilidad en el espacio: esto implica diferencias asociadas a la localización, las pueden estar dadas por, costo de transporte, las amenidades (externalidades asociadas a la ubicación) y diferencias en costos y beneficios sociales (paquetes fiscales y gastos que enfrentan en cada zona).

La vivienda se ve fuertemente influenciada por el desarrollo urbano. En este sentido la economía urbana ha demostrado una estrecha relación entre el crecimiento urbano y el precio de la vivienda. La escasez de suelo que se va generando a través del desarrollo de las ciudades, trae como consecuencia que los terrenos disponibles y aptos para la construcción residencial sean más reducidos y por lo tanto aumenten las disponibilidades de precios de dichos terrenos. Sin embargo la teoría ha demostrado que esta alza en los precios se compensa por mantenerse dentro de un radio de fácil acceso a los centros de consumo que supone una disminución en los costos de transporte, para el abastecimiento de bienes económicos.

-Durabilidad: Por la gran cantidad de años de vida útil, la vivienda no solo es un bien de consumo, sino también de inversión. Debido a su gran durabilidad, la oferta y demanda de viviendas deben responder a condiciones presentes y futuras de mercado.

La oferta es proporcionada principalmente por los stocks acumulables, originando una demanda de vivienda inelástica. Como consecuencia de esta inelasticidad, los ajustes en el precio se atribuyen principalmente a efectos de precios, los cuales ajustan variaciones en la demanda.

-Importancia presupuestaria: el gasto en vivienda (alquiler y mantenimiento) representa un gran porcentaje en los gastos totales de los individuos, aun más, en el caso de quienes alquilan servicios de vivienda. Por lo tanto el ingreso real disponible se convierte en una variable fundamental en la demanda de servicios de vivienda y por lo tanto en su precio.

-Fuertes asimetrías de información: Se puede producir grandes problemas de selección adversa al no conocer mucha de las características de las casas. En sí, lo que sucede es que los compradores solo conocen la función de distribución de las casas, pero no saben si la casa están enfrentando (buena o mal). Ante esto, lo que están dispuestos a pagar los compradores es un precio que surge como promedio de las casas buenas y malas.

-Costo de transacción importante: existen altos costos de búsqueda, monetarios y psicológicos que producen que la elasticidad precio de la demanda sea más reducida que a largo plazo¹.

-Necesidades básicas: La vivienda es un bien de consumo de primera necesidad que genera un importante interés social por el mercado de vivienda lo cual a su vez hace que la demanda sea más inelástica.

Enfoque teórico

El enfoque de los precios hedónicos reconoce que la vivienda es un paquete de características (antigüedad, cantidad de piezas, metros cuadrados, distancia de los servicios de comercios, etc.) donde no existe un mercado para la vivienda, sino uno implícito para cada característica de la misma. Entonces el precio de la vivienda surge de la suma de los precios de cada característica.

¹ La tasa de impaciencia para el consumo de vivienda es alta, por lo tanto, se está dispuesto a pagar un precio mayor.

precios implícitos. El hecho de que no exista un mercado para la vivienda, desde esta perspectiva, se debe a la existencia de múltiples características que diferencian una casa de otra. Ante esto no podemos homogeneizar las viviendas en un único mercado.

El objetivo principal del trabajo es estimar una función de precios hedónicos sobre las características asociadas a la construcción y localización de las viviendas.

Para justificar este objetivo empírico se utilizara un modelo teórico simple (Rosen) basado en los siguientes supuestos:

- La vivienda es un bien heterogéneo
- Hay un único consumidor
- Ignora el bien y se concentra en sus características. Es decir, lo que se compra son las características de la casa (localidad, antigüedad, etc.). Entonces el consumidor enfrenta un precio de la vivienda dado por la suma de los precios implícitos de las características de la vivienda.

$$P(z_1, z_2, \dots, z_k) = P(z)$$

Donde "p" es una función de precios hedónicos y los Z representan las características que definen esta función

Problema de las familias.

$$\text{Max}_{x,z} U(Z; X)$$

s/ a

$$P(Z) + X \leq Y$$

Z: es el vector de características de la vivienda.

X: es el bien numerario.

Y: es el ingreso.

Problema de las firmas

$$\text{Max} \pi = \mu \cdot P(Z) - C(\mu, Z)$$

μ : Cantidad de firmas iguales.

C.P.O para las familias

$$\frac{\partial U}{\partial Z_j} - \lambda \frac{\partial P}{\partial Z} = 0 \quad \Rightarrow \quad TMS_{Z_j, X} = \frac{\partial P}{\partial Z}$$

$$\frac{\partial U}{\partial X} - \lambda = 0$$

La sustitución entre el bien numerario y la vivienda es igual al cambio en el precio de la vivienda cuando se agrega una característica más.

Esta condición, nos dice que lo que las familias están dispuestas a pagar por una característica más (por ejemplo un baño más) debe igualar al valor en el que se incrementa el precio por esa característica adicional.

C.P.O de las firmas

$$\mu \frac{\partial P}{\partial Z} - \frac{\partial C}{\partial Z} = 0$$

$$\frac{\partial P}{\partial Z} = \frac{\frac{\partial C}{\partial Z}}{\mu}$$

Esta condición nos dice que el costo marginal de producción por unidad de producción iguala al valor que el mercado paga por ese producto.

Igualando las condiciones de las familias y de las firmas se obtiene la condición de equilibrio

$$TMS_{z,x} = \frac{\partial P}{\partial Z} = \frac{\frac{\partial C}{\partial Z}}{\mu}$$

Que nos dice que este se produce (el equilibrio) cuando la tasa marginal de sustitución iguala a lo que los individuos están dispuestos a pagar por una característica más y, con el tiempo, al costo marginal de producción por unidad de esa característica.

En términos generales $P(z_1, z_2, \dots, z_k) = P(z)$ es la función de precios que coordina las decisiones de consumidores y productores, por lo tanto es importante estimar esta función, lo cual justifica el objetivo de este trabajo.

IV

Parte empírica (estimación e inferencia)

La estimación de la función de precios hedónicos se hará a partir de la base de datos del paper *THERE GOES THE NEIGHBORHOOD, URBAN QUALITY OF LIFE AND REAL ESTATE PRICES IN BUENOS AIRES*, CRUCES, FERNÁNDEZ Y HAM (2008).

Se utiliza una función logarítmica-lineal (semilogarítmica)

$$\log \text{precio} = \alpha_1 + \alpha_2 \text{bathrooms} + \alpha_3 \text{dormitorio} + \alpha_4 \text{garage} + \alpha_5 \text{pisos} + \alpha_6 \text{dis_centrp} + \varepsilon$$

-La variable explicada está representada por el logaritmo del precio (*lnprecio*) .

-Las variables explicativas son las siguientes,

1. bathrooms: Cantidad de baños
2. dormitorio: Número de Dormitorios
3. garage: variable Dummy que toma el valor 1 si tiene garage y 0 si no lo tiene
4. pisos: Cantidad de pisos
5. dis_centrp: distancia al centro (en kms)

6. ε : es un término de error, el cual contiene a las variables explicativas no tenidas en cuenta en el modelo

Resultados de la regresión

. regress logprecio bathrooms dormitorio garage pisos dis_centrp

Source	SS	df	MS			
Model	1333.11743	5	266.623486	Number of obs =	5821	
Residual	1195.47728	5815	.205585087	F(5, 5815) =	1296.90	
Total	2528.59471	5820	.434466445	Prob > F	= 0.0000	
				R-squared	= 0.5272	
				Adj R-squared	= 0.5268	
				Root MSE	= .45341	

logprecio	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
bathrooms	.1629364	.0062005	26.28	0.000	.1507811	.1750917
dormitorio	.0642133	.0053935	11.91	0.000	.05364	.0747866
garage	.2334557	.0209068	11.17	0.000	.1924706	.2744408
pisos	.1401836	.010275	13.64	0.000	.1200408	.1603263
dis_centrp	-.0289439	.0018011	-16.07	0.000	-.0324748	-.0254131
_cons	11.21327	.0164871	680.12	0.000	11.18095	11.24559

Debido a que es una función semilogarítmica todos los coeficientes estimados se interpretan como el cambio porcentual en el log precio cuando aumenta (o disminuye) en una unidad la característica i .

El "R-square" es 0.5272 indicando que la variabilidad de las variables explicativas es una proporción igual al 52.72% de la variabilidad total de la variable explicada (Inpretación: quiere decir que existen otras variables explicativas que no fueron tenidas en cuenta en el modelo, las cuales están inmersas en el término de error y, por lo tanto, explican la proporción restante de la variabilidad total de la variable dependiente).

Podemos ver que todas las variables son significativas y positivas, excepto dis_centrp que es significativa y negativa, los valores observados por parte de sus estadísticos t son altos, lo que nos lleva a rechazar la hipótesis nula de insignificancia en la explicación del componente de la variable explicada. Esto quiere decir que las variables tenidas en cuenta en el modelo mantienen una relación lineal positiva con la variable dependiente.

La significatividad de cada una de las variables se mide realizando un test de hipótesis

$H_0: \alpha_i = 0$ Indica que no existe relación lineal entre log precio y la variable explicativa i .

$H_1: \alpha_i \neq 0$ Indica que existe relación entre log precio y la variable explicativa i .

$i = 2, \dots, 6$

El estadístico de prueba es el siguiente: $t = \frac{\hat{\beta}_i}{se(\hat{\beta}_i)} \approx t_{(n-k)}$

Ante esto, nuestra regla de decisión será rechazar H_0 si el valor "t" observado es mayor que el valor crítico (1.96) a un nivel de significación del 5%.
 Dicho de otra forma si el valor "p" es menor al nivel de significación del 5%, rechazamos H_0 . Tal como vemos en la regresión el valor "t" observado es mayor que el valor crítico en todos los casos, lo que nos lleva a obtener un valor p menor al nivel de significación, por lo tanto rechazamos la hipótesis nula argumentando que todas las variables explicativas introducidas en el modelo son relevantes para explicar el comportamiento de la variable "lnprecio".

Un intervalo de confianza contribuye a confirmar lo dicho, si mirando la salida de `summary()` esta nos indica que existe un 95% de confianza de que el coeficiente α_i no se encuentra fuera de ese intervalo. Ahora, como vemos, este intervalo [95% Conf. Interval] no posee el valor 0 en ninguno de los casos, lo cual nos proporciona información a favor del rechazo de H_0 . Observando el test de significatividad global F vemos que este toma un valor alto lo que muestra que las variables son significativas en forma conjunta..

Teniendo en cuenta el rechazo de la hipótesis nula de no significatividad podríamos concluir que no existe una alta multicolinealidad entre las variables explicativas. El supuesto de no multicolinealidad nos dice que no existe relación lineal entre las variables explicativas, supuesto se mantiene, los coeficientes estimados serían lineales, insesgados y, sobre todo, significativos. Igualmente llevaremos a cabo un test para afirmar nuestra conclusión.

```
corr logprecio pisos dis_centrp bathrooms dormitorio garage
(obs=5821)

-----+-----
      | logpre~o   pisos dis_ce~p bathro~s dormit~o   garage
-----+-----
logprecio | 1.0000
      pisos | 0.6299 1.0000
dis_centrp | 0.0961 0.3523 1.0000
      bathrooms | 0.6599 0.7021 0.2297 1.0000
dormitorio | 0.5347 0.6113 0.2119 0.5948 1.0000
      garage | 0.5582 0.7660 0.3908 0.6121 0.4784 1.0000
```

Lo que podemos ver es que existe cierto grado de correlación entre las variables explicativas, pero el mismo es muy bajo para generar que el estimador de la varianza de α_i sea alto, lo que nos llevaría a aceptar H_0 .

La alta multicolinealidad, puede generar que el estadístico "t" de significatividad individual sea chico cuando el R-square es aun alto. Esto nos puede conducir a aceptar H_0 ,

que por medio del estadístico "F" (de significatividad global) la estamos rechazando este podría arrojar un valor muy bajo. Cosa que no sucede en el modelo planteado.

La heterocedasticidad se produce cuando los términos de error dejan de tener la misma varianza, en ese caso los estimadores MCO no dejan de ser insesgados pero vuelven ineficientes, por lo tanto dejan de ser MELI.

El siguiente test sirve para evaluar si existe heterocedasticidad. Esta prueba utiliza los residuales obtenidos de la regresión original, los eleva al cuadrado y posteriormente los divide por las variables independientes. De este segundo modelo se obtiene el

estadístico $\frac{1}{2}SCE \approx X^2_{(p-1)}$. el cual se conoce como la prueba Breusch-Pagan

distribuye como una chi-cuadrado con p-1 grados de libertad donde "p" es el número de regresores en la regresión auxiliar.

Las hipótesis se formulan suponiendo que la varianza del error se describe con una función lineal de las variables Z no estocásticas $\sigma_i^2 = \alpha_1 + \alpha_2 Z_{2i} + \dots + \alpha_m Z_{mi}$

Si $\alpha_2 = \alpha_3 = \dots = \alpha_m = 0$ (hipótesis nula), entonces $\sigma_i^2 = \alpha_1$ lo cual es una constante indica la ausencia de heterocedasticidad por el contrario, si $\alpha_2 \neq 0 \vee \alpha_3 \neq 0 \vee \dots$ (hipótesis alternativa) estaríamos en presencia de heterocedasticidad.

La regresión de los residuos al cuadrado respecto a las variables explicativas es la siguiente

```
. . reg errcuadrados bathrooms dormitorio garage pisos dis_centrp
```

Source	SS	df	MS	
Model	47.1582023	5	9.43164046	Number of obs = 5821
Residual	1221.8621	5815	.21012246	F(5, 5815) = 44.89
Total	1269.02031	5820	.218044726	Prob > F = 0.0000
				R-squared = 0.0372
				Adj R-squared = 0.0363
				Root MSE = .45839

errcuadrados	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
bathrooms	.0345649	.0062686	5.51	0.000	.0222762 .0468536
dormitorio	.0119381	.0054527	2.19	0.029	.0012488 .0226275
garage	-.0365792	.0211362	-1.73	0.084	-.0780141 .0048557
pisos	.0139715	.0103877	1.35	0.179	-.0063923 .0343354
dis_centrp	-.0220895	.0018209	-12.13	0.000	-.0256591 -.0185199
_cons	.2954049	.016668	17.72	0.000	.2627293 .3280804

$$\frac{1}{2}SCE \approx X^2_{(p-1)} = 23.579101 \approx X^2_4$$

El valor crítico a un nivel de 95 % de confianza es 9.48 este dado que el observado es 23.57 se rechaza la hipótesis nula de que los errores son homocedásticos. Esto provoca que los estimadores dejen de ser MELI.

Otro test utilizado es el de White el cual plantea que para detectar la heterocedasticidad en el modelo, se debe realizar una regresión sobre los errores al cuadrado y analizar la siguiente prueba; se toma el R^2 de la regresión sobre los errores al cuadrado y se multiplica por el número de observaciones, asumiendo que tiene una distribución X^2 se somete a la siguiente prueba: si el estadístico observado es mayor al valor crítico, entonces se rechaza la hipótesis nula, es decir, el modelo es heterocedástico. El valor obtenido es 216.54 el cual es mayor al valor crítico de 9.48, por lo tanto se rechaza el resultado del test anterior el cual afirma la presencia de heterocedasticidad.

Una forma de solucionar la heterocedasticidad es a través del método de estimación por mínimos cuadrados generalizados (MCG) el cual consiste en transformar las variables originales para que estas satisfagan los supuestos del modelo clásico y luego aplicar la estimación de MCO sobre el modelo transformado, de esta forma se asegura que los estimadores obtenidos sean MELI.

Teóricamente

$$y_i = \beta_1 X_{0i} + \beta_2 X_i + \mu_i$$

Suponiendo que se conocen las varianzas heterocedásticas σ_i dividiendo las variables del modelo por la varianza obtenemos el modelo transformado

$$y_i \frac{1}{\sigma_i} = \beta_1 X_{0i} \frac{1}{\sigma_i} + \beta_2 X_i \frac{1}{\sigma_i} + \mu_i \frac{1}{\sigma_i}$$

A partir de esta transformación comprobamos que este modelo presenta errores homocedásticos, lo cual asegura que si aplicamos el método de MCO al modelo transformado se obtendrán como ya se dijo, estimadores que sean MELI.

$$\text{Var}\left(\mu_i \frac{1}{\sigma_i}\right) = E\left(\mu_i \frac{1}{\sigma}\right)^2 = \frac{1}{\sigma_i^2} E(\mu_i^2) = \frac{1}{\sigma_i^2} (\sigma_i^2) = 1$$

$$\text{Var}\left(\mu_i \frac{1}{\sigma_i}\right) = 1$$

Lo cual es una constante, por lo tanto la varianza del termino de error transformado es homocedastica.

Otra alternativa es estimar el modelo original por MCO pero computar estimadores de las varianzas de los coeficientes, para esto se utiliza el estimador de White que correctamente la segunda columna de la salida de regresión

```
regress logprecio dis_centrp pisos garage dormitorio bathrooms, robust
```

	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
dis_centrp	-.0289439	.0018612	-15.55	0.000	-.0325926	-.0252953
pisos	.1401836	.0127179	11.02	0.000	.1152518	.1651153
garage	.2334557	.0207651	11.24	0.000	.1927484	.274163
dormitorio	.0642133	.0074096	8.67	0.000	.0496876	.078739
bathrooms	.1629364	.0082791	19.68	0.000	.1467062	.1791666
_cons	11.21327	.0190962	587.20	0.000	11.17583	11.2507

Los coeficientes obtenidos en esta regresión son los mismos de la regresión original.

La ausencia de Autocorrelacion supone que el término de error asociado a una observación no está influenciado por el término de error asociado a cualquier otra observación, en términos formales $E(\mu_i, \mu_j) = 0$, por lo tanto la presencia de autocorrelacion puede expresarse de esta forma $E(\mu_i, \mu_j) \neq 0$.

La correlación serial o autocorrelacion puede ocurrir por diversas razones

- En el caso de series de tiempo, puede ocurrir que observaciones sucesivas sean interdependientes
- Sesgo de especificación debido a variables omitidas
- Forma funcional incorrecta del modelo
- Cuando se utilizan rezagos en las series de tiempo, entre otros.

Forma de detectar la autocorrelacion

El estadístico Durbin-Watson es el siguiente

$$d = \frac{\sum_{t=2}^{t=N} (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{t=N} \varepsilon_t^2}$$

Trabajando analíticamente se llega a $d = 2(1 - \rho)$

Siendo ρ una estimación del coeficiente de correlación

Como $-1 \leq \rho \leq 1$ esto implica que $0 \leq d \leq 4$ cualquier valor estimado de d debe estar entre estos valores, teniendo en cuenta los valores que toma ρ si se encuentra que d es igual a 2 se puede suponer que no existe correlación de primer orden, si $\rho = +1$ indica la presencia de una perfecta correlación positiva en los residuos $d = 0$. Por lo tanto cuanto más cerca de 0 este d mayor será la evidencia de una correlación serial positiva.

Si $\rho = -1$ entonces $d = 4$, entonces cuanto más cerca de 4 este el estadístico, más fuerte es la evidencia de una correlación serial negativa, también hay que tener en cuenta que este test también arroja valores en los cuales no se puede decidir si aceptar o rechazar la hipótesis nula, las cuales son llamadas zonas de indecisión.

Este test se utiliza generalmente para datos de series de tiempo por lo tanto no se utilizó en este trabajo.

Octubre de 2008

Bibliografía

Cruces, Fernandez y Ham (2008). "There goes the neighborhood, urban quality of life and real estate prices in Buenos Aires.

Gujarati, "Econometría". McGraw Hill.

Rosen, S (1974). "Hedonic Prices and Implicit Markets: Product Differentiation under **Perfect Competition**"; Journal of Political Economy.