

El análisis de regresión múltiple y su aplicación al estudio de las rentas diferenciales del suelo

José Miguel SANTOS PRECIADO

Dos de los componentes que intervienen de manera fundamental en el precio de la vivienda son los costes derivados de la construcción y los del suelo urbanizado que se necesita disponer para edificar. La proporción de estos elementos entre sí, así como respecto al total de gastos que la fabricación de la mercancía vivienda conlleva, depende de un conjunto de factores, estructurales y coyunturales, difíciles de evaluar. La situación técnica del sector de la construcción, la calidad y características de las viviendas a construir y la disponibilidad del suelo urbanizable barato en el mercado del suelo son, entre otros, factores esenciales para explicar el costo total de la vivienda.

A. Santillana (1972) ha estudiado para Madrid, con datos de 1965, la relación existente entre los costos debidos al suelo urbanizado y los correspondientes a la construcción de la vivienda (ver tabla número 1), obteniendo valores que varían entre 1,55 para las viviendas de alta calidad y 0,55 en las viviendas de calidad inferior. La repercusión del suelo parece ser más definitiva en las viviendas destinadas a las rentas superiores que en las más bajas.

La explicación de esta desproporción sería debido, según el autor, a las diferencias de precio de los terrenos, consecuencia de los siguientes tipos de rentas:

- Renta de situación, que aparece debido a las diferencias de gastos de transporte según el emplazamiento del terreno y que, en general, guarda alguna relación con la distancia al centro.
- Renta de zona o barrio, ligada a la segregación social del espacio, lo que hace que determinadas personas estén dispuestas a

TABLA 1
COSTO DE LA VIVIENDA

	<i>Alta calidad</i>	<i>Clase media</i>	<i>Clase inferior</i>
Suelo urbanizado	40,0	25,0	20,0
Coste construcción	25,8	34,8	36,0
Beneficio constructor	3,3	4,5	4,6
Beneficio promotor	20,0	26,5	30,7
Gastos generales	10,9	9,2	8,7
	100,0	100,0	100,0

pagar más para vivir con los de su grupo social. En la mayoría de las ocasiones, estas zonas poseen mejores características de calidad de paisaje y comunicaciones.

- Rentas derivadas de una monopolización de suelo (que aumenta los precios de la oferta) o bien de ausencia de éste, por no haber sido preparado para construir, y que produce las mismas consecuencias.
- Rentas de volumen, derivadas del mayor o menor índice de edificabilidad permitido legalmente en la construcción.

El interés de nuestro estudio se centra en obtener las rentas diferenciales del suelo, existentes en los municipios del A. M. M. en el año 1978, intentando deducir cuál de las causas antes expuestas pueden explicarlas. Para conseguirlo, hemos utilizado el modelo matemático contenido en el análisis de regresión múltiple, que pasamos brevemente a exponer.

1. EL MODELO DE REGRESIÓN LINEAL MÚLTIPLE: EL SIGNIFICADO DE LOS RESIDUALES

Este modelo de regresión múltiple supone la relación directa lineal entre los valores adoptados por una variable dependiente (Y) y los de un conjunto de variables independiente (X_1, X_2, \dots, X_n), mediante una ecuación del tipo

$$Y = a + b_1 X_1 + b_2 X_2 + \dots + b_n X_n$$

siendo a, b_1, b_2, \dots, b_n constantes, que representan el término independiente de la ecuación y los pesos proporcionales atribuidos a cada una

de las variables. X_1, X_2, \dots y X_n , en la participación funcional de obtención de la variable Y .

Los valores adoptados por la variable Y en un conjunto N de casos observados (en nuestro caso promociones de viviendas) son, lógicamente, diferentes entre sí. ($Y_1, Y_2, \dots Y_n$) y distintos a su valor medio \bar{Y} . Estas diferencias pueden ser explicadas, en parte, por los distintos valores obtenidos en los N lugares por las variables X_1, X_2, \dots y X_n , si suponemos la existencia de una relación funcional entre ellas y la variable dependiente Y , según la ecuación de regresión. Como la ecuación no se cumplirá totalmente, existirán unas diferencias entre los valores teóricamente previstos ($Y'_1, Y'_2, \dots Y'_n$) y los realmente obtenidos (Y_1, Y_2, Y_n), llamadas residuales. En la figura núm. 1 expresamos el proceso antes descrito:

<u>Variables</u>	<u>Valores de las variables en los N casos</u>
Y	$Y_1, Y_2 \dots\dots\dots Y_N$ $\bar{Y} = \frac{\Sigma Y_i}{N}$
X_1	$X_{11}, X_{12} \dots\dots\dots X_{1N}$
.	
.	
.	
X_n	$X_{n1}, X_{n2} \dots\dots\dots X_{nN}$
<hr/>	
<u>Valor esperado matemáticamente</u>	
Y'	$Y'_1, Y'_2 \dots\dots\dots Y'_N$
<hr/>	
<u>Residual</u>	
$Y - Y'$	$Y_1 - Y'_1, Y_2 - Y'_2, \dots\dots\dots Y_N - Y'_N$

¿Cómo podremos medir el acercamiento de la realidad al modelo? Se demuestra matemáticamente que el error cuadrático medio, res-

pecto a su media, adoptado por la variable Y, puede descomponerse en dos sumandos:

$$\frac{\sum (Y - \bar{Y})^2}{N} = \frac{\sum (Y - Y')^2}{N} + \frac{\sum (Y' - \bar{Y})^2}{N}$$

¿Qué significa cada uno de estos términos?

$\frac{\sum (Y - \bar{Y})^2}{N}$ significa una medida de la variación de los valores que adapte la variable Y respecto a su valor medio.

$\frac{\sum (Y' - \bar{Y})^2}{N}$ significa lo mismo, pero de los valores teóricos Y'.

$\frac{\sum (Y - Y')^2}{N}$ es una forma de medir el valor de los residuos o diferenciales, entre los valores reales y teóricos de la variable Y.

Es lógico que cuanto mayor sea el segundo término respecto al tercero, mayor validez tendrá la regresión analizada, pues un porcentaje mayor de la variedad (varianza) de la variable Y (representado por el primer término) será explicado por las variables X_1, X_2, \dots y X_n y no por motivos imprevistos.

Otra forma de medir el valor de la regresión es a partir del coeficiente de correlación múltiple de Y sobre X_1, X_2, \dots, X_n . Este coeficiente representa el coeficiente de correlación de Pearson, de los valores adoptados por la variable Y y los teóricos correspondientes a Y'. Se demuestra matemáticamente la siguiente relación entre este coeficiente y los tres términos analizados con anterioridad.

$$R^2_{Y, X_1, X_2, \dots, X_n} = 1 - \frac{\sum (Y - Y')^2}{\sum (Y - \bar{Y})^2} = \frac{\sum (Y' - \bar{Y})^2}{\sum (Y - \bar{Y})^2}$$

Por lo tanto, el cuadrado del coeficiente de correlación múltiple representa el tanto por uno de la varianza asociado a la regresión respecto a la total.

$$\text{Si } R_{Y, X_1, X_2, \dots, X_n} = 0,9 \text{ y } R^2_{Y, X_1, X_2, \dots, X_n} = 0,81$$

significaría que el 81 por 100 de las diferencias individuales de Y son explicadas por las variaciones de X_1, X_2, \dots y X_n y sólo el 19 por 100 se debería a causas no previstas.

Además, es posible efectuar un test de significación del análisis de regresión realizado. Esta inferencia estadística se realiza a partir de suponer una serie de presupuestos que deben cumplir los residuales (esperanza matemática nula, normalidad, varianza constante, independencia). Se demuestra entonces que la relación:

$$\frac{\frac{\sum (Y' - \bar{Y})^2}{n - 1}}{\frac{\sum (Y - Y')^2}{N - n}}$$

da lugar a una distribución F con $n - 1$ y $N - n$ grados de libertad.

Si en nuestro caso, el valor calculado para F supera el valor de la tabla para un porcentaje de significación elevado, supone que el análisis de regresión es satisfactorio por salirse ampliamente de la situación aleatoria.

2. AMPLIACIÓN DEL MODELO DE REGRESIÓN AL CÁLCULO DE LAS RENTAS DIFERENCIALES DEL SUELO

Ya indicamos, previamente a la exposición del modelo de regresión múltiple, que el costo de la vivienda puede ser explicado en gran parte por medio de sus dos componentes principales; costo del suelo y costo de la construcción. Nuestra hipótesis de trabajo consiste en suponer que el costo de la edificación debe corresponderse con una serie de características y servicios (superficie, calefacción, ajardinado, garaje, juegos infantiles, piscina..., etc.), más fácilmente medibles, estableciendo una relación entre el resto del precio de la vivienda hasta completar el total¹ y las rentas residuales derivadas del suelo. Para ello, pretendemos aplicar el modelo de regresión lineal múltiple, utilizando como variables independientes una serie de variables relacionadas con las características de la vivienda y como variable dependiente el precio de la misma. Los valores residuales que se obtengan deben reflejar en gran parte el costo del suelo.

Hemos utilizado para el análisis los datos obtenidos por D. A. T. I. N.² sobre promociones inmobiliarias en los municipios del área

¹ En el costo de la vivienda intervienen también los conceptos de beneficios del constructor y promotor y los gastos generales. El valor residual integraría también estos costos que no tenemos en cuenta. A pesar del error que cometemos, continúa existiendo una relación grande entre el valor residual y el costo del suelo.

² Informe sobre el mercado inmobiliario DATIN, COPLACO, Madrid, 1978.

metropolitana madrileña, durante el período 1974-78. Los datos hacen referencia a las características de 115 promociones de viviendas situadas en los municipios de Colmenar Viejo, Alcobendas, San Sebastián de los Reyes, La Moraleja³, Coslada, San Fernando de Henares, Torrejón de Ardoz, Alcalá, Arganda, Getafe, Pinto, Valdemoro, Parla, Fuenlabrada, Leganés, Alcorcón, Móstoles, Pozuelo, Majadahonda y Las Rozas. El número de promociones analizadas es proporcional al de viviendas construidas en cada municipio:

Las variables utilizadas como independientes han sido:

- 1) Calefacción.
- 2) Juegos infantiles.
- 3) Ajardinamiento.
- 4) Piscina.
- 5) Aparcamiento.
- 6) Superficie.

Para su medición se ha establecido el siguiente baremo:

— *Calefacción.*

Calefacción eléctrica: 1.

Calefacción gas: 2.

Calefacción central: 3.

— *Juegos infantiles, ajardinamiento y piscina.*

No existencia: 1.

Existencia: 2.

— *Aparcamiento.*

No existencia: 1.

Aparcamiento en el exterior: 2.

Aparcamiento en el edificio: 3.

— *Superficie.*

La superficie está expresada en Dm.².

³ La Moraleja es la única unidad de las analizadas que no comprende un término municipal completo y como tal la reflejamos.

Hemos realizado un primer análisis de regresión lineal múltiple, utilizando como variable dependiente el precio total de la vivienda y como variables independientes las expresadas con anterioridad. Los resultados pueden observarse en la tabla núm. 2. El cuadrado del coeficiente de regresión múltiple es 0,7779, lo que indica que la regresión explica un 77,79 por 100 de la varianza existente, con un valor del test F igual a 63,055, muy elevado y que confirma en principio la validez del análisis. Sin embargo, un estudio detallado de la contribución de cada una de las variables a la explicación de la varianza total pone de manifiesto que sólo dos variables son significativas (superficie y calefacción) y que una sola variable (la superficie) contribuye en el 75,48 por 100 de la explicación. Esta desproporción hay que buscarla en la gran importancia que la superficie posee en el precio total de la vivienda.

TABLA 2
ANALISIS REGRESION MULTIPLE NUM. 1

Variable dependiente:	Precio total vivienda	
Variables independientes:	1) Superficie	R = 0,8820
	2) Calefacción	R ² = 0,7779
	3) Ajardinamiento	F = 63,055
	4) Juegos infantiles	Porcentaje de varianza explicada 77,79 %.
	5) Piscina	
	6) Aparcamiento	

Ecuación de regresión:

$$y = - 29,91 + 45,77 X_1 + 2,62 X_2 + 1,71 X_3 + 0,75 X_4 - 0,55 X_5 + 2,18 X_6$$

Introducción de variables:

	R	R ²	F	Total porcentaje Varianza
— 1) Superficie	0,8688	0,7548	347,80	75,48
— 2) Calefacción	0,8763	0,7679	6,33	76,79
— 3) Resto	No significativas		4	77,79

Para obviar esta dificultad hemos creído conveniente utilizar como variable dependiente el precio de la vivienda por unidad de superficie, realizando de nuevo el análisis, con los resultados de la tabla número 3. El porcentaje de varianza explicada es ahora inferior, 41,93 por 100, y el valor del test F igual a 12,995, es suficientemente significativo. El porcentaje, que puede parecer reducido, no lo es tanto, si pensamos que el resto de la varianza no explicada se debe a los valores diferenciales del suelo, que pueden presentar variaciones muy fuertes según la localización de las promociones. La participación de las variables independientes es más equilibrada, existiendo tres variables significativas (aparcamiento, superficie y calefacción).

TABLA 3
ANALISIS DE REGRESION MULTIPLE NUM. 2

Variable dependiente:	Precio vivienda/m ²	
Variables independientes:	1) Superficie	R = 0,6475
	2) Calefacción	R ² = 0,4193
	3) Ajardinamiento	F = 12,995
	4) Juegos infantiles	Porcentaje de varianza
	5) Piscina	explicado 41,93 %
	6) Aparcamiento	

Ecuación de regresión:

$$Y = 7,80 + 8,45 X_1 + 2,59 X_2 + 0,82 X_3 + 1,15 X_4 - 1,76 X_5 + 2,87 X_6$$

Introducción de variables:

	R	R ²	F	Total porcentaje Varianza
— 1) Aparcamiento	0,5558	0,3089	50,51	30,89
— 2) Superficie	0,6108	0,3730	11,44	37,30
— 3) Calefacción	0,6427	0,4130	7,56	41,30
— 4) Resto	0,6475	0,4193	4	41,93

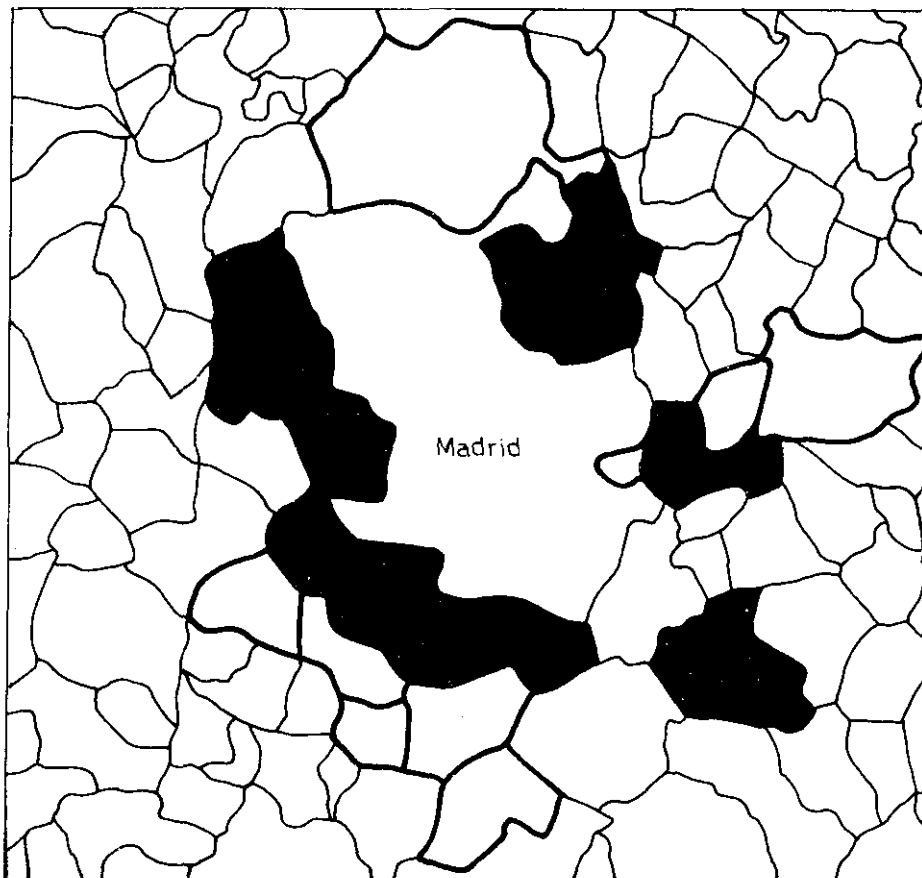


GRÁFICO 1.—Municipios valor residual positivo.

Los valores residuales medios por municipios vienen recogidos en la tabla núm. 4, y representados gráficamente en el gráfico núm. 1. Un valor positivo de los residuales significa que el precio pagado por cada metro cuadrado de vivienda excede al que teóricamente le correspondería según la ecuación de regresión. Deducimos que estas diferencias son debidas al mayor costo de la vivienda procedente de las rentas diferenciales del suelo, cuyas causas queremos deducir en los municipios afectados.

Hemos realizado un análisis de regresión lineal simple entre los valores residuales medios obtenidos por municipio y los precios medios de las viviendas, obteniendo una correlación muy elevada ($r = 0.876$), como se pone de manifiesto en el gráfico núm. 2. Estos

TABLA 4
VALORES RESIDUALES MEDIOS POR MUNICIPIOS

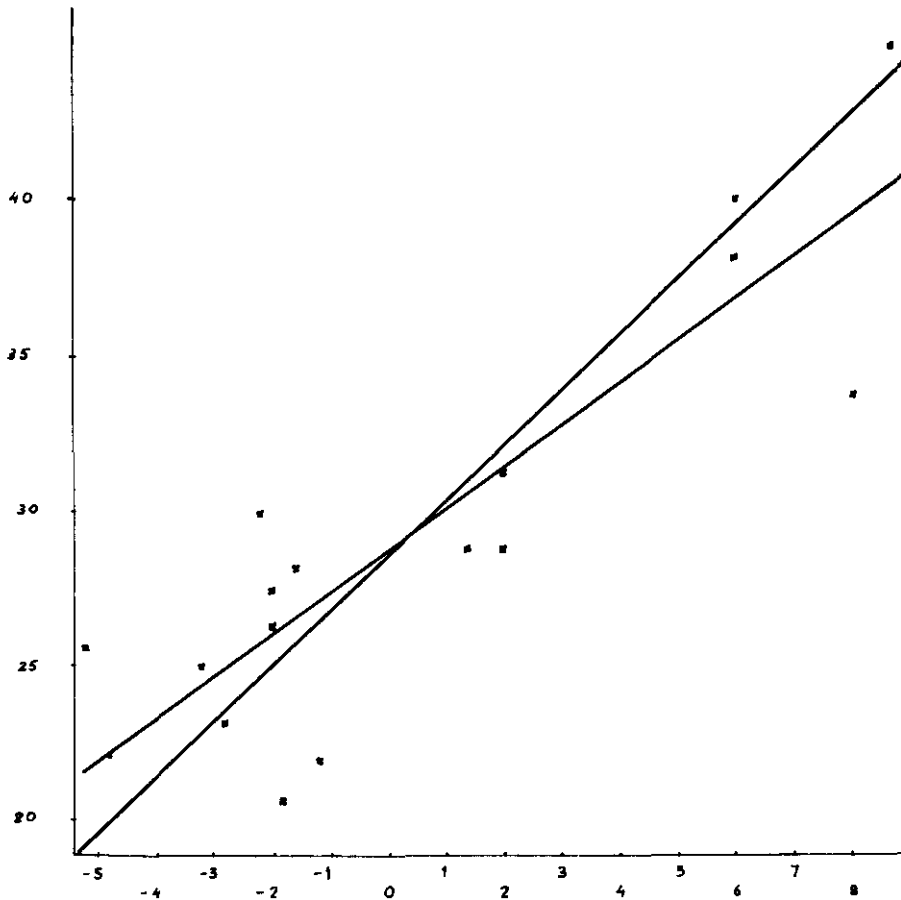
1. Colmenar Viejo	- 2,1737
2. Alcobendas	4,1431
3. San Sebastián de los Reyes	0,6089
4. La Moraleja	8,5521
5. Coslada	- 1,6317
6. San Fernando de Henares	3,8960
7. Torrejón de Ardoz	- 2,0673
8. Alcalá de Henares	- 2,0929
9. Arganda	- 2,7237
10. Getafe	8,0418
11. Pinto	- 5,0965
12. Valdemoro	- 5,2294
13. Parla	- 1,8226
14. Fuenlabrada	- 1,2113
15. Leganés	1,9504
16. Alcorcón	1,3719
17. Móstoles	- 3,2946
18. Pozuelo de Alarcón	5,9100
19. Majadahonda	3,8500
20. Las Rozas	12,7196

resultados corroboran la hipótesis de trabajo de partida, en el sentido de que los costos del suelo tienen una repercusión mayor en aquellos municipios de superior nivel de renta de sus habitantes y de precio de vivienda más elevado.

Los municipios representados en el gráfico núm. 1 pueden agruparse en tres grupos, cuyas características homogéneas reflejan el distinto valor de sus valores residuales medios.

Estos grupos serían:

1) Municipios de la zona Oeste y Norte constituidos por la Moraleja, Majadahonda, Pozuelo y Las Rozas, cuyo valor positivo de la renta diferencial del suelo puede ser explicado por la segregación social de sus moradores con niveles de renta más elevados.



COR = 0. 876

$$X = 0.49815 \cdot Y - 14.426$$

$$Y = 1.5393 \cdot X + 29.140$$

GRÁFICO 2.—Regresión residuales-precio vivienda.

2) Municipios situados en la corona de proximidad de Madrid, como Alcorcón, Getafe, San Fernando de Henares, Alcobendas y San Sebastián de los Reyes, cuyos valores del suelo se deben al aumento de precio debido a la proximidad a la capital.

3) Municipios más alejados de Madrid, en la zona Sur y Este, como Móstoles, Fuenlabrada, Parla, Pinto, Valdemoro, Arganda, Torrejón de Ardoz y Alcalá de Henares, donde las disponibilidades del suelo son superiores, produciendo residuales negativos.

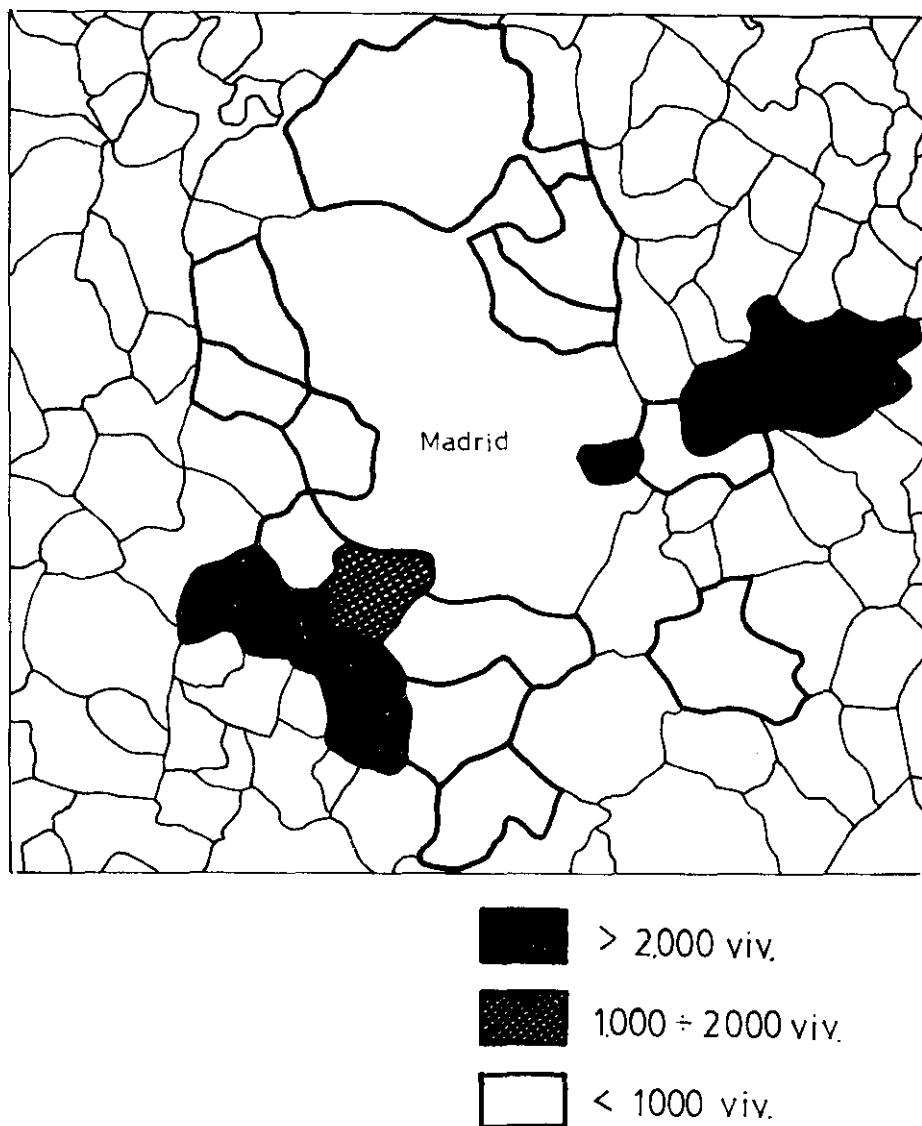


GRÁFICO 3.—Oferta de viviendas.

Por último, hemos querido analizar dónde se concentra la oferta de vivienda en el A. M. M., comparándola con los valores residuales obtenidos en el análisis de regresión múltiple efectuado. Los valores se reflejan en la tabla núm. 5 y los gráficos núms. 3 y 4. En el gráfico número 3 puede observarse que la oferta de viviendas se ha desplazado

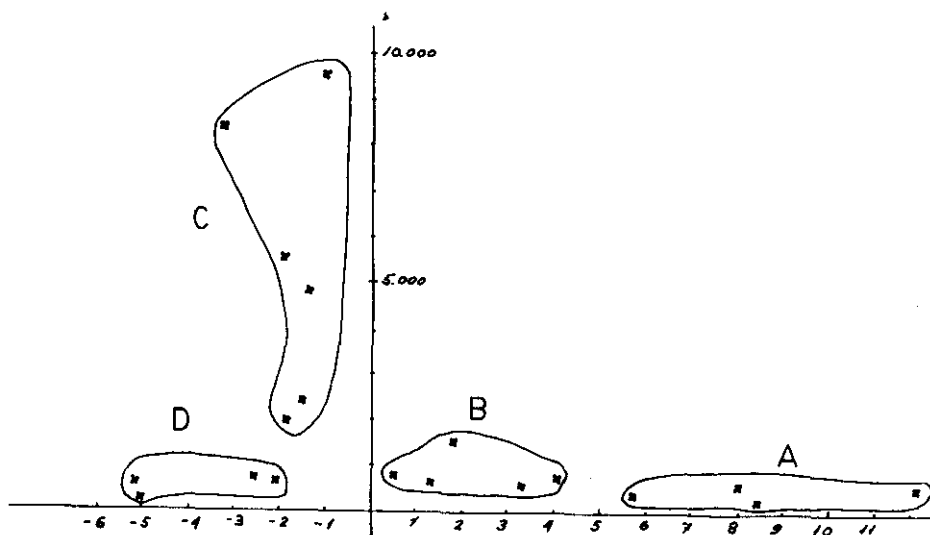


GRÁFICO 4.—Residuales-viviendas ofertadas.

TABLA 5
TOTAL VIVIENDAS CONSTRUIDAS

1. Colmenar Viejo	647
2-3. Alcobendas - S. Sebastián de los R. ...	1.604
4. La Moraleja	226
5. Coslada	4.814
6. San Fernando de Henares ⁴	—
7. Torrejón de Ardoz	2.001
8. Alcalá	5.487
9. Arganda	686
10. Getafe	761
11. Pinto	152
12. Valdemoro	745
13. Parla	2.375
14. Fuenlabrada	9.610
15. Leganés	1.382
16. Alcorcón	588
17. Móstoles	8.824
18. Pozuelo	354
19-20. Majadahonda - Las Rozas	1.322

⁴ No existen datos.

hacia los municipios del exterior, donde los costos del suelo son inferiores.

En el gráfico núm. 4 pueden apreciarse la formación de cuatro grupos de municipios de características similares. Los grupos A y B, con residuales positivos y una reducida oferta de viviendas, presentan una cierta correlación negativa entre estas dos variables. El grupo A (de mayores rentas diferenciales de suelo) coincide en general con los municipios de mayor nivel de renta (a excepción de Getafe). El grupo C, de mayor oferta de viviendas, está compuesto por los municipios del Sur y del Este, correspondientes en gran parte a la segunda corona del área metropolitana, hacia donde se ha desplazado el crecimiento metropolitano. Existe una correlación positiva entre las mayores ofertas de viviendas y las rentas diferenciales de suelo más reducidas. Al final, el grupo D está integrado por cuatro municipios (Arganda, Colmenar Viejo, Pinto y Valdemoro), en donde los residuales son negativos y la oferta de viviendas es reducida, que corresponde a los municipios más alejados y de menor desarrollo.

Diciembre, 1982

BIBLIOGRAFIA

Santillana del Barrio, Antonio (1972): *Análisis económico del problema de la vivienda*. Ed. Ariel, Barcelona.

RESUMEN

El crecimiento de la ciudad metropolitana produce una diferenciación de los precios de la vivienda en los municipios de la periferia. Esta variación está determinada por la suma de sus tres componentes básicos: el precio del suelo, el coste de la construcción y gestión y el beneficio del promotor. El empleo del modelo de regresión lineal múltiple a un conjunto de variables, relacionadas con la calidad de la vivienda, sirve para deducir de su costo los valores correspondientes a las rentas diferenciales del suelo.

ABSTRACT

The growth of the metropolitan city produces a differentiation in housing prices in the peripheric municipalities. This variation is determined by the sum of the three basic components: the price of the ground, the cost of the construction and management and the producer's profit. The use of the multiple lineal regression model in a whole of variables, related to the housing quality, serves for deducting the corresponding values to the differential ground rents from its cost.

RÉSUMÉ

La croissance de la ville métropolitaine produit une différenciation dans les prix de logement aux municipalités périphériques. Cette variation s'établit en additionnant les composantes élémentaires: prix du sol, le coût de la construction et gestion et le profit du promoteur. L'emploi du modèle de régression lineal multiple sur un groupe de variables en rapport avec la qualité de logement sert pour réduire de son coût les valeurs correspondants aux rentes différentielles du sol.