

## **SEMINARIO ECONOMETRIA 14 DE MAYO DEL 2012**

### **CASO: Determinando un estándar de productividad comercial en banca**

**1.1. Objetivo del análisis:** Un banco quiere estimar la productividad estándar esperada de cada equipo comercial disponible en cada una de sus oficinas. El objetivo de esta estimación es que sirva para tomar las correspondientes decisiones sobre el dimensionamiento (ó tamaño) de los equipos comerciales y la redistribución de recursos entre oficinas.

Usando un modelo de regresión múltiple, los pasos del análisis son:

- a) Determinar, para cada equipo comercial considerado, una productividad estándar esperada en función de las características del equipo y de los recursos de que dispone.
- b) Definir la productividad esperada a partir de un modelo estadístico de la productividad observada, ajustado a partir de una muestra de equipos comerciales comparables.
- c) Fundamentar las decisiones que se tomen en las diferencias existentes entre la productividad observada y la productividad esperada.

**1.2. Datos disponibles:** El banco dispone de la siguiente información sobre 554 de sus oficinas o sucursales.

**TIPO\_OFICINA:** Variable discreta que toma valor 1 si la oficina es urbana y cero si la oficina es rústica.

**COMPLEJIDAD:** Variable acotada entre 1 y 7. El valor 1 se corresponde con el nivel mínimo de complejidad de la oficina y el 7 con el máximo. Puesto que es una información cualitativa (más que cuantitativa) de la oficina, a partir de la misma se construyen 7 variables ficticias, en donde  $C_i$  ( $i=1, 2, \dots, 7$ ) es una variable que toma el valor 1 si la oficina tiene un grado de complejidad  $i$ -ésimo y cero, en cualquier otro caso.

**TOTAL\_CLTES:** Número de clientes de la oficina. La oficina con menos clientes tiene 300 y el máximo se corresponde con una oficina de 21.000 clientes.

**FUERZA\_VENTAS:** Número de personas que componen el equipo comercial en cada oficina. El equipo más pequeño es de 2 personas y el máximo lo componen 13 personas.

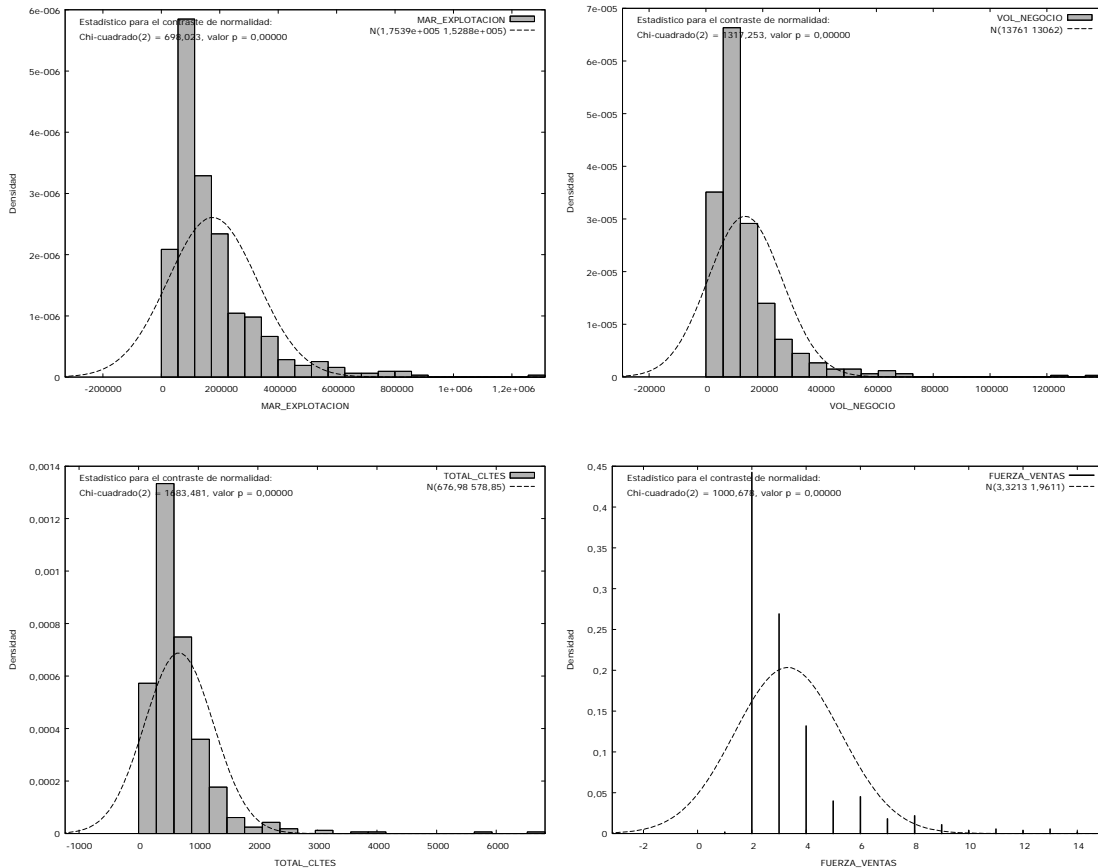
**VOL\_NEGOCIO:** Variable que define el volumen del negocio de la oficina a través de la suma de depósitos más préstamos.

**MAR\_EXPLOTA:** Define el margen de explotación ordinario de la oficina. Esta es la variable que representa la productividad observada de una oficina, dadas sus características, algunas de ellas observadas dadas las variables anteriores y otras no disponibles o no observables.

La unidad de medida de las variables **VOL\_NEGOCIO** y **MAR\_EXPLOTA** es monetaria, pero estos datos se han escalado por motivos de confidencialidad.

**1.3. Características de los datos**: Antes de especificar una relación entre el margen de explotación y las variables que miden complejidad, el tipo de la oficina, el volumen de negocio y de clientes y el tamaño del equipo, conviene “inspeccionar” algunas características de los datos gráficamente.

**Gráficos 1.3.1: Distribuciones de probabilidad de MAR\_EXPLOTA, VOL\_NEGOCIO, TOTAL\_CLTES y F\_VENTAS frente a la distribución normal**

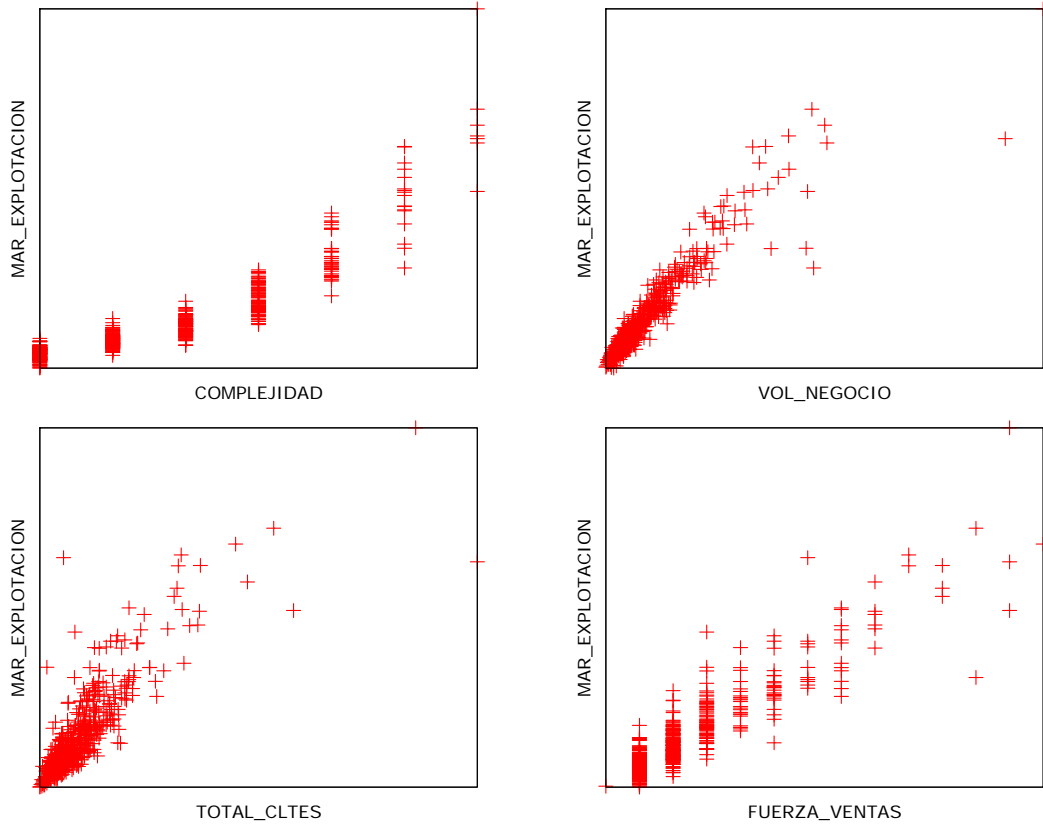


En los gráficos 1.3.1 se observa que, en general, las variables no se distribuyen según un modelo de probabilidad normal. En todos los casos presentados, se observa una distribución asimétrica y un exceso de curtosis. Como ejemplo, la variable con un mayor coeficiente de asimetría es la de **TOTAL\_CLTES** con un valor de 4.49 y con un exceso de curtosis de 33.20. La variable de **FUERZA\_VENTAS** presenta el menor coeficiente de asimetría (2.41) y también el mínimo exceso de curtosis (6.94).

Las relaciones entre pares de variables (la endógena versus una variable explicativa) se muestran en los gráficos 1.3.2. de la página siguiente. Es fácil apreciar que la relación entre el **MAR\_EXPLOTA** y la **COMPLEJIDAD** de la oficina no es lineal. Además, la relación entre el **MAR\_EXPLOTA** y el **VOL\_NEGOCIO** y el **TOTAL\_CLTES** muestra que la mayoría de las oficinas de esta muestra tienen pocos clientes, poco volumen de negocio y como consecuencia, poco margen de

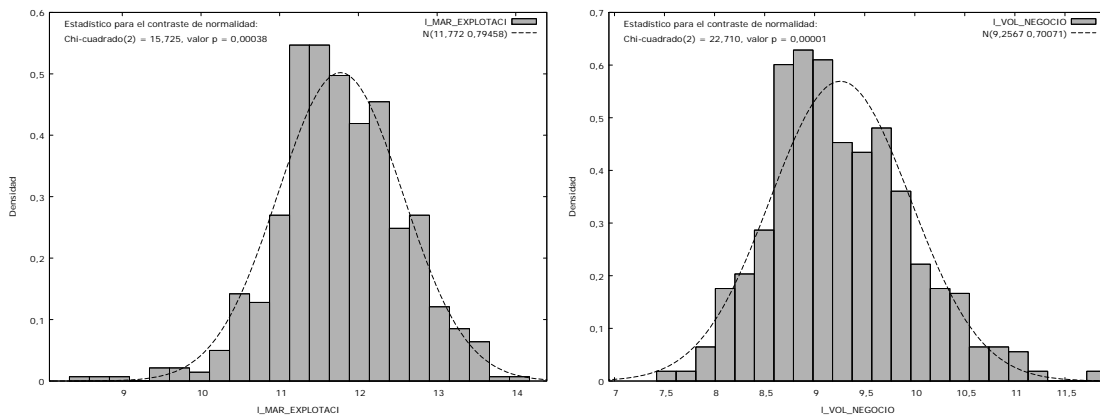
explotación. A pesar de ser un grupo mayoritario el de las oficinas pequeñas, se observa que tienen poco peso relativo para “anclar” la recta de regresión.

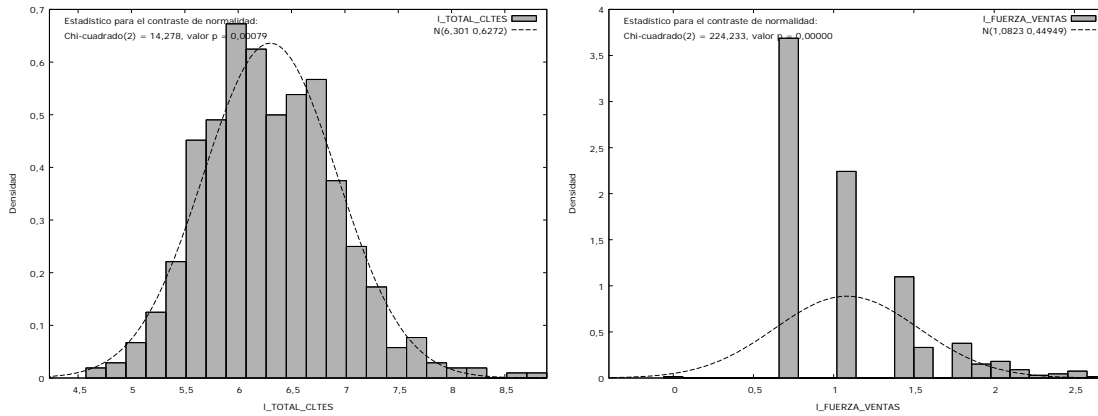
**Gráficos 1.3.2. Relaciones en el plano X-Y entre pares de variables (en el eje Y siempre se representa el MAR\_EXPLOTAION)**



Al tomar logaritmos, algunos de estos problemas se resuelven (por ejemplo, se induce normalidad en los datos). Es sabido que la transformación logarítmica comprime los valores altos de una variable y expande los valores bajos.

**Gráficos 1.3.3. Distribuciones de probabilidad del logaritmo de las variables MAR\_EXPLOTA, VOL\_NEGOCIO, TOTAL\_ CLTES y F\_VENTAS frente a la normal.**





Como se aprecia en los gráficos 1.3.3, la transformación logarítmica hace que los datos de esta muestra de oficinas sea más homogénea que en nivel. Los coeficientes de asimetría y los excesos de curtosis se reducen de forma importante, aunque formalmente el contraste de normalidad de Jarque-Bera sigue rechazando normalidad al 5%.

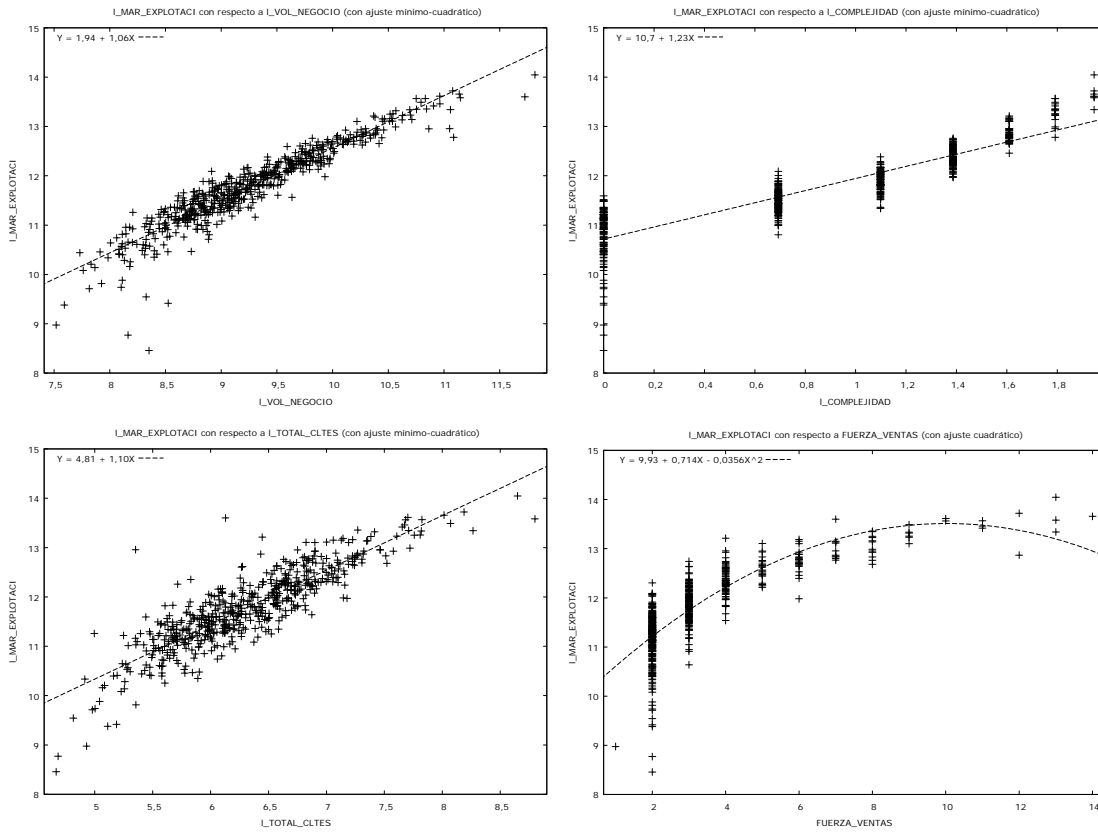
Variable	Coefficiente de asimetría	Exceso de curtosis
L_MAR_EXPLORACION	-0.25	0.85
L_VOL_NEGOCIO	0.47	0.19
L_TOTAL_CLTES	0.40	0.48
L_FUERZA_VENTAS	1.09	0.72

De hecho, la variable menos log-normal es la de fuerza de ventas, que aún después de la transformación presenta una asimetría positiva elevada (1.09).

Las relaciones entre variables en logaritmos, (ver Gráficos 1.3.4) son mucho más lineales que en nivel, salvo en el caso de la variable FUERZA\_VENTAS, en donde observa que a partir de un equipo de aproximadamente 8 personas, una persona más en el equipo ya no supone un aumento en el margen de la oficina (es decir, se puede pensar en una relación cuadrática entre L\_MAR\_EXPLORACION y la FUERZA\_VENTAS).

Sin embargo, incluso especificando una relación cuadrática entre ambas variables, se aprecia visualmente que las oficinas con equipos comerciales más pequeños (y también con menos clientes, poca complejidad y poco volumen de negocio) serán las peor ajustadas (explicadas) por un modelo de regresión lineal.

### Gráficos 1.3.4. Relaciones en el plano X-Y entre variables en log (excepto para la variable FUERZA\_VENTAS)



### 1.4. Estimación de un modelo inicial

A continuación, se presenta la estimación MCO de un modelo inicial, en donde se utiliza la muestra total, es decir, las 554 oficinas. La complejidad de las mismas se incluye como factor explicativo a través de las 6 variables ficticias, eliminando el nivel de complejidad mínimo (C1), para evitar multicolinealidad exacta al tener un término constante.

La estimación del modelo 1, presentado en la página siguiente, sugiere los siguientes comentarios:

- 1) La bondad del ajuste es del 90.15% en esta preliminar estimación. Es decir, más del 90% de la variabilidad del margen de explotación de la oficina viene explicado por las variables explicativas incluidas. Esto sin hacer muchos esfuerzos en la modelización (forma funcional, tratamiento de atípicos, etc.)
- 2) A pesar del alto valor del  $R^2$ , no son variables individualmente significativas a la hora de explicar el margen ni la  $FUERZA\_VENTAS$ , ni los niveles de complejidad intermedios y altos (desde el nivel 3 hasta el nivel máximo).

3) El signo negativo del coeficiente asociado al tipo de oficina, indica que el margen esperado (en log) de una oficina urbana sería menor que el asociado a un oficina rústica. Es decir, la diferencia en el margen (en log) de una oficina rústica con respecto a una urbana, con las mismas características de negocio, número de clientes, complejidad, etc., es positiva.

4) El signo negativo del coeficiente de la fuerza de ventas es contraintuitivo, pero dicho coeficiente no es estadísticamente significativo ( $p\_valor = 0.38$ ).

5) La variable más relevante, tanto desde el punto de vista estadístico como cuantitativo, es el volumen de negocio (en log), por encima del total de clientes.

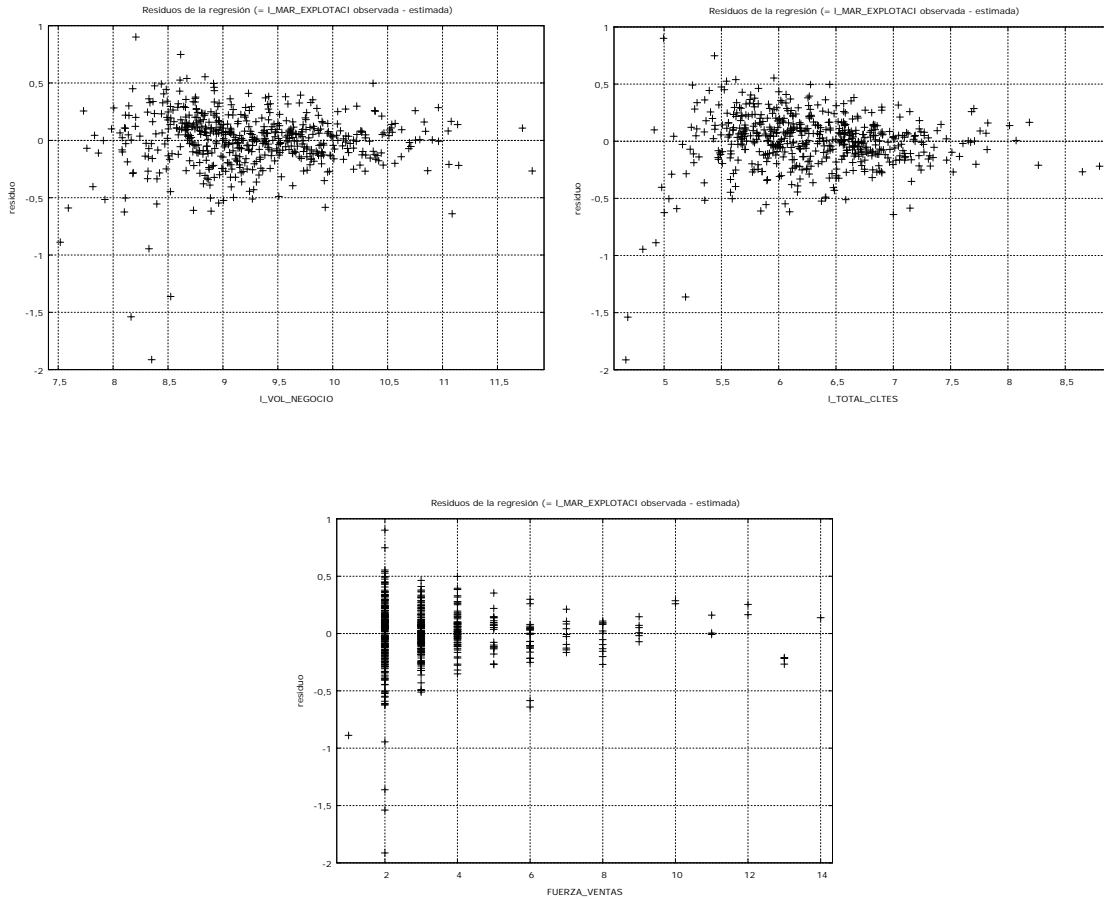
**Modelo 1: MCO, usando las observaciones 1-554**  
**Variable dependiente: I\_MAR\_EXPLOTA**

	<i>Coeficiente</i>	<i>Desv. Típica</i>	<i>Estadístico t</i>	<i>Valor p</i>	
Const	1,93601	0,475799	4,0690	0,00005	***
I_VOL_NEGOCI	0,840661	0,0524439	16,0297	<0,00001	***
I_TOTAL_CLTES	0,328661	0,0410414	8,0080	<0,00001	***
FUERZA_VENTA	<b>-0,014085</b>	0,0159636	-0,8823	<b>0,37799</b>	
C2	0,128793	0,0431357	2,9858	0,00296	***
C3	0,101281	0,0627179	1,6149	0,10692	
C4	0,109203	0,0848792	1,2866	0,19879	
C5	0,0944552	0,121304	0,7787	0,43652	
C6	0,0407741	0,161989	0,2517	0,80136	
C7	-0,121527	0,218211	-0,5569	0,57781	
T_OFICINA	<b>-0,0932604</b>	0,0228122	-4,0882	<b>0,00005</b>	***

Media de la vble. dep.	11,77202	D.T. de la vble. dep.	0,794577
Suma de cuad. residuos	34,36690	D.T. de la regresión	0,251577
<b>R-cuadrado</b>	<b>0,901566</b>	R-cuadrado corregido	0,899754
F(10, 543)	497,3405	Valor p (de F)	6,8e-266
Log-verosimilitud	-16,01233	Criterio de Akaike	54,02466
Criterio de Schwarz	101,5135	Crit. de Hannan-Quinn	72,57661

Con el objetivo de chequear este modelo inicial, se recuperan los residuos resultantes del modelo 1 y se dibujan versus el comportamiento de las variables explicativas continuas (todas salvo las ficticias de complejidad y de tipo de oficina).

**Gráficos 1.4.1. Gráficos de dispersión de los residuos del modelo 1 con respecto a las variables de tamaño en log (volumen de negocio, volumen de clientes y tamaño del equipo)**



Tanto los gráficos anteriores como el contraste formal de White muestran el rechazo de la ausencia de heteroscedasticidad en los residuos resultantes del modelo 1. Visualmente se observa mayor varianza en los residuos para las oficinas pequeñas (con poco negocio, pocos clientes y equipos comerciales pequeños) que para las oficinas medianas y grandes. El valor del estadístico de contraste de White es de 252.7 con un  $p\_valor$  prácticamente nulo.

El hecho de que las oficinas más pequeñas y con menos negocio sean las peor ajustadas por el modelo, puede provocar heteroscedasticidad. No obstante, la solución más adecuada sería segmentar la muestra o bien, eliminar aquellas oficinas atípicas por ser excesivamente pequeñas.

Se reestima el modelo inicial corrigiendo heteroscedasticidad bajo la fórmula de White. De esta forma, al menos, estarán correctamente calculadas las desviaciones típicas de los parámetros y será válidos los contrastes de hipótesis que hagamos con el nuevo modelo (modelo 2).

**Modelo 2: con corrección de heterocedasticidad, usando las observaciones 1-554**  
**Variable dependiente: I\_MAR\_EXPLOTA CI**

	<i>Coficiente</i>	<i>Desv. Típica</i>	<i>Estadístico t</i>	<i>Valor p</i>	
Const	4,69561	0,441169	10,6436	<0,00001	***
I_VOL_NEGOCI	0,652235	0,0456278	14,2947	<0,00001	***
I_TOTAL_CLTES	0,120963	0,0340485	3,5527	0,00041	***
FUERZA_VENTA	0,0248041	0,00970912	2,5547	0,01090	**
C2	0,234987	0,0400537	5,8668	<0,00001	***
C3	0,32308	0,0527627	6,1233	<0,00001	***
C4	0,452185	0,0729271	6,2005	<0,00001	***
C5	0,496112	0,0982565	5,0492	<0,00001	***
C6	0,554481	0,123767	4,4800	<0,00001	***
C7	0,54192	0,183725	2,9496	0,00332	***
T_OFICINA	-0,104623	0,0163917	-6,3827	<0,00001	***

**Estadísticos basados en los datos ponderados:**

Suma de cuad. residuos	2158,461	D.T. de la regresión	1,993757
R-cuadrado	0,935848	R-cuadrado corregido	0,934667
F(10, 543)	792,1311	Valor p (de F)	0,000000
Log-verosimilitud	-1162,808	Criterio de Akaike	2347,616
Criterio de Schwarz	2395,105	Crit. de Hannan-Quinn	2366,168

**Estadísticos basados en los datos originales:**

Media de la vble. dep.	11,77202	D.T. de la vble. dep.	0,794577
Suma de cuad. residuos	38,10755	D.T. de la regresión	0,264914

Las conclusiones del modelo 2 estimado son:

- Todos los coeficientes son estadísticamente significativos, al menos, al 5% de significación. El p\_valor más alto se corresponde con la FUERZA\_VENTAS, siendo igual a 0.0109. Por tanto, ésta es la variable menos significativa estadísticamente a la hora de explicar el margen.
- El ajuste pasa a ser del 93.58%. Es decir, sin mucho esfuerzo, el ajuste dentro de la muestra sigue siendo muy elevado.
- Para interpretar los resultados numéricos del modelo 2, lo más fácil es fijar un valor base. Por ejemplo, el margen esperado de explotación para una oficina urbana de nivel de complejidad 2, con un volumen de negocio de 8000 unidades monetarias, 500 clientes y un fuerza de ventas de 2 personas (el mínimo), se calcula como:

$$e^{4.69561} \times 8000^{0.652235} \times 2^{0.0248041} \times e^{0.0234987} \times e^{-0.104623}$$

El margen de explotación esperado para una oficina con las mismas características, pero:

- Con un 1% más de volumen de negocio, aumentaría en un 0.65%
- Con un 1% más de clientes, aumentaría en un 0.12%



- Con una persona más en la fuerza de ventas, se incrementaría en un 2.48%
- Con un aumento de complejidad desde el nivel 1 al 2, el margen esperado aumentaría en un 23.50%; desde el nivel 1 al 3, aumentaría en un 32.31%; desde el nivel 1 al 4 en un 45.22%; del nivel 1 al 5 en un 49.61%; del nivel 1 al 6 en un 55.45% y del nivel mínimo al máximo en un 54.19%.
- Desde el punto de vista cuantitativo, se observa gran similitud entre los coeficientes asociados a niveles de complejidad cercanos (el nivel 2 y 3, o bien, el nivel 6 y 7). Este resultado sugiere que se podría simplificar el modelo incorporando la restricción de la igualdad de los coeficientes asociados a distintos niveles de complejidad.
- Finalmente, por término medio, una oficina situada en zona rústica es más rentable que una situada en zona urbana. En concreto, la diferencia en el margen de explotación de una oficina rústica con respecto a una urbana es del 10.46%.

### **1.5. Residuos multiplicativos y usos del modelo**

Los residuos del modelo 2 son desviaciones entre el margen de explotación observado y el esperado (en log). Es fácil convertir los residuos en errores multiplicativos y dibujar su distribución muestral (ver gráfico 1.5.1.).

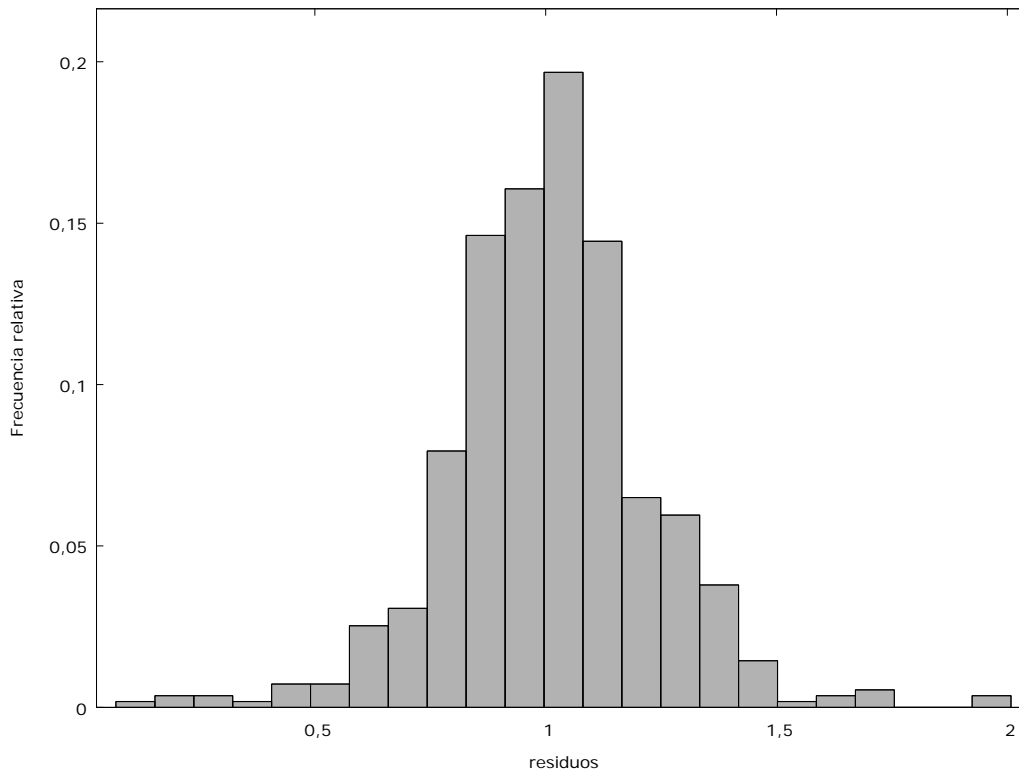
La interpretación de esta figura es la siguiente:

- Un residuo multiplicativo igual a uno implica que el margen de explotación esperado para esa oficina coincide con el margen observado. En estos casos, no habría que actuar.
- Un residuo multiplicativo igual a 0.5 implica que el margen observado en esa oficina es un 50% menor que el esperado, dadas sus características.
- Un residuo multiplicativo igual a 1.5 supone que, en este caso, el margen observado es un 50% superior al esperado.

La importancia de la distribución de estos residuos es su utilidad para definir una política de redistribución de recursos a partir de:

- Una serie de intervalos o de puntos de corte que se deciden a priori.
- Una política específica para las oficinas que caigan en cada intervalo definido. Por ejemplo, redistribuir recursos desde las oficinas que están en el intervalo (1.4 – 1.8) hacia las que están en el intervalo (0.2 – 0.4).

**Gráfico 1.5.1. Distribución muestral de los residuos multiplicativos MCO resultantes del modelo 2**



**1.6. Simplificando el modelo**

Si escribimos el modelo teórico correspondiente al modelo 2 estimado:

$$l\_MAR\_EXPLOTACI_i = \beta_0 + \beta_1 l\_VOL\_NEGOCIO_i + \beta_2 l\_TOTAL\_CLTES_i + \beta_3 FUERZA\_VENTAS_i + \alpha_2 C2_i + \alpha_3 C3_i + \alpha_4 C4_i + \alpha_5 C5_i + \alpha_6 C6_i + \alpha_7 C7_i + \beta_4 T\_OFICINA + \varepsilon_i$$

Dado que la estimación de los coeficientes asociados a las variables de nivel de complejidad son muy similares para niveles de complejidad parecidos (C2 y C3 ó C4 y C5) , una forma de simplificar el modelo 2 es contrastar las siguientes hipótesis simples:

Restricción:  
 $b[C2] - b[C3] = 0$

Estadístico de contraste:  $F(1, 543) = 7,49785$ , con valor  $p = 0,00638001$

Restricción:  
 $b[C4] - b[C5] = 0$

Estadístico de contraste:  $F(1, 543) = 1,16487$ , con valor  $p = 0,280936$

donde  $b[Ci]$  ( $i=2,3,4,5$ ) es el coeficiente asociado a la variable que aparece en el corchete. Dado el  $p\_valor$  de ambos contrastes, se rechaza la hipótesis de que el

coeficiente asociado al nivel de complejidad 2 coincide con el del asociado al nivel 3, mientras que no se puede rechazar a los niveles de significación habituales (5% y el 10%), la igualdad de los coeficientes asociados a los niveles de complejidad 4 y 5. Los resultados del modelo estimado incorporando dicha restricción, se ofrecen a continuación (ver Modelo 3). Los criterios de información de Akaike ó de Schwarz también sugieren como preferible el modelo 3 al modelo 2.

**Modelo 3: con corrección de heterocedasticidad, usando las observaciones 1-554**  
**Variable dependiente: I\_MAR\_EXPLOTA**

	<i>Coefficiente</i>	<i>Desv. Típica</i>	<i>Estadístico t</i>	<i>Valor p</i>	
const	4,57726	0,39144	11,6934	<0,00001	***
I_VOL_NEGOCI	0,654596	0,0402991	16,2434	<0,00001	***
I_TOTAL_CLTES	0,135001	0,0339785	3,9731	0,00008	***
FUERZA_VENTA	0,0341124	0,00893991	3,8157	0,00015	***
C2	0,224064	0,0379386	5,9060	<0,00001	***
C3	0,299921	0,0469285	6,3910	<0,00001	***
<b>C45</b>	<b>0,421036</b>	<b>0,0657279</b>	<b>6,4057</b>	<b>&lt;0,00001</b>	<b>***</b>
C6	0,453947	0,094436	4,8069	<0,00001	***
C7	0,413951	0,155847	2,6561	0,00814	***
T_OFICINA	-0,105248	0,0166596	-6,3176	<0,00001	***

**Estadísticos basados en los datos ponderados:**

Suma de cuad. residuos	2117,899	D.T. de la regresión	1,973118
R-cuadrado	0,937173	R-cuadrado corregido	0,936134
F(9, 544)	901,6372	Valor p (de F)	0,000000
Log-verosimilitud	-1157,553	Criterio de Akaike	2335,106
Criterio de Schwarz	2378,278	Crit. de Hannan-Quinn	2351,972

**Estadísticos basados en los datos originales:**

Media de la vble. dep.	11,77202	D.T. de la vble. dep.	0,794577
Suma de cuad. residuos	37,88827	D.T. de la regresión	0,263908

**1.7. Actividades obligatorias para el estudiante a partir del CASO**

- Dada la relación entre I\_MAR\_EXPLOTA y la FUERZA\_VENTAS que se observa en el gráfico 1.3.4. ¿cómo reformularía el modelo 2 estimado?
- ¿Es significativa la diferencia en el margen de explotación esperado en las oficinas con niveles de complejidad 6 y 7 en el modelo 3? Si no lo fuera ¿cómo incluiría esta información en el mismo?
- ¿Es posible simplificar aún más el modelo estimado en el apartado b, redefiniendo el nivel de complejidad de la oficina?
- Elimine las oficinas más pequeñas de la muestra (no más de un 20%) y reestime el modelo 1 por MCO.
- Contraste heteroscedasticidad gráfica y formalmente en el modelo estimado en el apartado d. Si es necesario corregir del problema, hágalo.
- ¿Cambiaría las recomendaciones derivadas del apartado 1.5 con el nuevo modelo estimado en el apartado e?

