

TRABAJO FIN DE GRADO

CURSO 2015-2016

**ESTIMACIÓN DE LA FUNCIÓN DE PRODUCCIÓN DEL SECTOR AGRARIO EN
LA UNIÓN EUROPEA: UN ESTUDIO ECONOMETRICO**



Autor: Sergio Sánchez Zapata

Tutor: Juan Aparicio Baeza

DNI: 45561137T

Fecha de entrega: 30 de noviembre de 2015

Grado de Administración y Dirección de Empresa

Facultad Ciencias Sociales y Jurídicas de Orihuela

Universidad Miguel Hernández de Elche

Contenido

RESUMEN	5
INTRODUCCIÓN	6
BREVE INTRODUCCIÓN A LA UE	6
POLÍTICA AGRARIA DE LA UE	8
OBJETIVOS	10
METODOLOGÍA.....	10
BASE DE DATOS	12
ESTUDIO ECONÓMTRICO	16
Estudio econométrico completo	16
Análisis preliminar.....	16
AJUSTE DEL MODELO	19
BONDAD DEL AJUSTE	21
Error estándar residual:	21
Tabla de Anova:.....	21
Coeficiente de determinación R^2 :	21
DIAGNÓSTICO DEL MODELO	22
Normalidad.....	23
Homocedasticidad.....	24
Incorrelación	25
Media cero	26
REDUCCIÓN DE VARIABLES	26
Método de significatividad estadística.....	28
ESTUDIO ECONÓMTRICO DEL MODELO SIMPLIFICADO	31
ANÁLISIS PRELIMINAR(Gráfico-Numérico).....	31
AJUSTE DEL MODELO	33
BONDAD DEL AJUSTE	34
Error estándar residual:	34
Tabla de Anova:.....	34
Coeficiente de determinación R^2 :.....	35
DIAGNÓSTICO DEL MODELO	35
Normalidad.....	35
Homocedasticidad.....	37
Incorrelación	38

Media cero	38
CONCLUSIONES	40
REFERENCIAS.....	42



RESUMEN

En este trabajo estudiamos una base de datos de agricultura europea con el fin último de estimar la función de producción que nos permita establecer una relación entre los factores productivos (inputs) y cantidad de producto (output) bajo cierta tecnología en el contexto del sector agrario. Para ello, tomamos como output la variable *Farm Net Value Added* (SE415), que recoge la remuneración asociada a los factores fijos de producción (trabajo, tierra y capital). Este indicador es sensible a los métodos de producción empleados. Como inputs, por otro lado, seleccionamos los insumos que mejor se relacionan con la variable respuesta, utilizando para ello registros oficiales de la Unión Europea.

El trabajo se inicia con una breve introducción a la Unión Europea y las políticas agrarias comunes para después comentar qué es una función de producción y, en concreto, la función de producción Cobb-Douglas. Las variables de la base de datos oficial utilizada serán transformadas mediante logaritmos para poder realizar un estudio econométrico completo de un modelo lineal. Una vez alcanzado el mejor modelo explicativo, establecemos las conclusiones del trabajo.

Según nuestros hallazgos, el valor añadido neto del sector agrario en la Unión Europea puede explicarse estadísticamente de forma adecuada a través de dos variables: por un lado, los insumos específicos de cosechar (semillas, fertilizantes, etc.) junto con el stock de inputs específicos y costes específicos de la silvicultura, y, por otro lado, la renta pagada por cultivar (edificios y costes de alquiler). Adicionalmente, se estima en el modelo que los rendimientos a escala de la tecnología imperante presentan una tendencia decreciente.

INTRODUCCIÓN

BREVE INTRODUCCIÓN A LA UE

La Unión Europea (UE) es una asociación económica y política integrada actualmente por 28 países del continente. El origen de la UE se encuentra en el periodo posterior a la Segunda Guerra Mundial. Sus primeros pasos consistieron en impulsar la cooperación económica con la idea de que, a medida que aumentara la interdependencia económica entre los países, disminuirían las posibilidades de conflicto.

En 1957 la UE fue fundada por seis estados: Alemania, Bélgica, Francia, Italia, Luxemburgo y los Países Bajos. La ampliación de la UE con la incorporación de nuevos países desde 1957 se ha incrementado de 6 a 28 estados miembros.

La tabla de datos con la que vamos a trabajar incluye los siguientes países de la unión europea:

Países analizados	Fecha de incorporación a la UE
(BEL) Belgium	1958
(BGR) Bulgaria	2007
(CYP) Cyprus	2004
(CZE) CzechRepublic	2004
(DAN) Denmark	1973
(DEU) Germany	1958
(ELL) Greece	1981
(ESP) Spain	1986
(EST) Estonia	2004
(FRA) France	1958
(HUN) Hungary	2004
(IRE) Ireland	1973
(ITA) Italy	1958
(LTU) Lithuania	2004
(LUX) Luxembourg	1958
(LVA) Latvia	2004
(MLT) Malta	2004
(NED) Netherlands	1958
(OST) Austria	1995
(POL) Poland	2004
(POR) Portugal	1986
(ROU) Romania	2007
(SUO) Finland	1995
(SVE) Sweden	1995
(SVK) Slovakia	2004
(SVN) Slovenia	2004
(UKI) UnitedKingdom	1973

Tabla 1. Relación de países analizados y sus fechas de incorporación a la UE. Fuente: Comisión Europea

Según el tratado de la Unión Europea la adhesión la puede solicitar cualquier estado europeo que respete los valores democráticos de la UE y se compromete a defenderlos.

Además también deben de cumplir otros criterios más específicos (Europea s.f.)¹:

Criterios políticos. Los países deben tener instituciones estables que garanticen la democracia, el estado de derecho y los derechos humanos.

Criterios económicos. Los países deben tener una economía de mercado en funcionamiento y la capacidad de afrontar la presión competitiva y las fuerzas de mercado dentro de la Unión. Criterios jurídicos. Los países deben de aceptar las normas y prácticas establecidas por la UE y, en particular, los grandes objetivos de unión política, economía y monetaria.

Lo que empezó como una unión económica es ahora una organización activa en todos los aspectos políticos, esto quedó reflejado en 1993 cuando se cambió el nombre de CEE (Comunidad Económica Europea) a UE (Unión Europea).

La UE se basa en el Estado de Derecho y todas sus actividades están fundadas en los tratados acordados voluntaria y democráticamente por todos los países miembros.

El mercado único permite que puedan circular libremente la mayoría de mercancías, servicios, personas y capitales y es el principal motor económico de la UE. El objetivo de este mercado único es desarrollarlo al máximo para que los europeos lo podamos aprovechar al máximo.

En lo referente a la política agrícola los países de la UE aceptaron centralizar plenamente sus competencias, además de la financiación pública necesaria. Por lo tanto, las decisiones y las ayudas no están en manos de cada país, sino que son responsabilidad de la UE en su conjunto.

¹ Comisión europea. http://ec.europa.eu/index_es.htm

POLÍTICA AGRARIA DE LA UE

La política agraria común (PAC) fue creada en 1962 y sus principales objetivos son (Europea s.f.)²:

- Ayuda a los agricultores para producir suficientes alimentos para Europa.
- Garantiza que los alimentos sean seguros.
- Protege a los agricultores de la excesiva volatilidad de los precios y de las crisis de mercado.
- Les ayuda a invertir en la modernización de sus explotaciones.
- Mantiene comunidades rurales viables, con economías diversificadas.
- Crea y mantiene puestos de trabajo en la industria alimentaria.
- Protege el medio ambiente y el bienestar de los animales.

La política agrícola de la UE ha evolucionado considerablemente en las últimas décadas para ayudar a los agricultores a hacer frente a nuevos retos y adaptarse a la evolución de las actitudes ciudadanas. Las sucesivas reformas han llevado a los agricultores a basar su producción en la demanda del mercado y no en las decisiones venidas de Bruselas.

A continuación repasamos los principales hitos de la PAC.

En 1984, debido a los éxitos de la PAC, las explotaciones son tan productivas que se producen excedentes y la PAC introduce medidas para ajustar los niveles de producción a las necesidades del mercado.

En 1992 la PAC desplaza su apoyo al mercado hacia el productor. El apoyo en los precios disminuye y se sustituye por pagos directos a los productores. También se les alienta a ser más respetuosos con el medio ambiente. Esta reforma coincide con la Cumbre de la Tierra de Río 1992, en la que hace su aparición el principio de desarrollo sostenible.

En 2003 la PAC proporciona ayudas a la renta suprimiendo las que van ligadas a la producción, a condición de cumplir una serie de normas que son la de atender las tierras agrícolas, las de medioambiente, la de bienestar animal y la seguridad de los alimentos.

Las reformas más recientes de 2013 dirigen la política agraria hacia prácticas agrarias más sostenibles, investigación y divulgación de los conocimientos, un sistema más justo de ayudas a los agricultores y una posición más fuerte de los agricultores en la cadena alimentaria.

Se busca reforzar la competitividad del sector agrícola promoviendo la agricultura sostenible y la innovación, apoyando el empleo y el crecimiento de las zonas rurales.

Otras metas importantes de la PAC son la ayuda a los consumidores a elegir sus alimentos con conocimiento de causa gracias a sistemas de etiquetado de calidad de la UE, fomentar la innovación en la agricultura y la en la transformación e impulsar relaciones comerciales justas con los países en desarrollo.

² Comisión Europea. Dirección General de comunicación. http://europa.eu/pol/index_es.htm

Los retos futuros de la PAC son por un lado la previsión de que la producción mundial de alimentos se duplicará en 2050 y las repercusiones del cambio climático.

Aproximadamente el 40% del presupuesto de la UE va destinado a la PAC pero se debe fundamentalmente a que es uno de los sectores que depende directamente de la UE, pero también hay que indicar que en los últimos 30 años ha bajado del 75% al 40%.



OBJETIVOS

El objetivo del presente trabajo es estimar la función de producción del sector agrario de la Unión Europea (Europea s.f.)³.

Tomando como variable respuesta la remuneración de los factores fijos de producción ("Farm Net ValueAdded") de la agricultura europea de una relación de 27 países de la UE estudiaremos su relación con las siguientes variables:

Total labour input (SE010): Input del trabajo total expresado en AWU (unidad de trabajo anual = equivalente al tiempo completo de un persona)

Total UtilisedAgriculturalArea (SE025): Consiste en la ocupación de la tierra en propiedad, alquilada y compartida (Se expresa en hectáreas 10.000 m2)

Total livestockunits (SE080): Media anual del número de equinos, ganadería, ovejas, cerdos y avícolas en stock.

Total specificcosts (SE281): Inputs específicos de cosechar (semillas, fertilizantes, etc..), stock de inputs específicos y costes específicos de la silvicultura.

Total farmingoverheads (SE336): Costes de abastecimiento relacionados con la producción pero no relacionados con líneas específicas de producción.

Rentpaid (SE375): Renta pagada para cultivar, por los edificios y los costes de alquiler.

Total fixedassets (SE441): Capital de la tierra cultivable, de los edificios y la silvicultura + Edificios +Maquinaria y equipo + el stock de cría

En el caso de que haya una relación satisfactoria entre la variable respuesta y alguna o todas las variables explicativas estimaremos la función de producción.

METODOLOGÍA

Una función de producción de una industria establece las relaciones técnicas existentes entre los factores productivos (inputs) y cantidad de producto (output) bajo cierta tecnología.

Una de las formas funcionales más utilizada en la literatura de producción en economía es la función Coob-Douglas (Aparicio 2004)⁴,

$$Y = A L^{\alpha} K^{\beta}; \quad \text{donde } 0 < \alpha, \beta < 1$$

Donde Y es la producción (output), L el factor trabajo (input 1) y K el factor capital (input 2). A, α y β son constantes positivas, donde A es un factor de escala y α y β representan las

³ http://ec.europa.eu/agriculture/rica/database/database_en.cfm

⁴ Aparicio, J., Martínez, M., y Morales, J. (2004). *Modelos Lineales Aplicados en R*. Universidad Miguel Hernández.

elasticidades del output a variaciones en las cantidades de los factores productivos, L y K, respectivamente.

La elasticidad total de producción (o elasticidad de escala), ϵ , mide el cambio proporcional en el output resultante de aumentar equiproporcionalmente todos los inputs. La elasticidad total de producción puede calcularse como la suma de las elasticidades parciales, es decir, con respecto a cada input. En el caso de la función de Cobb-Douglas: $\epsilon = \alpha + \beta$.

El valor de ϵ se encuentra relacionado con el retorno de escala de la tecnología de la siguiente manera:

Retorno de escala	Elasticidad Total de Producción (ϵ)
Constante	=1
Creciente	>1
Decreciente	<1

Tabla 2. Tabla de posibles valores de la elasticidad.

Podemos expresar la función de producción de Cobb-Douglas de forma econométrica como $Y_i = A L_i^\alpha K_i^\beta e^{u_i}$; donde u representa la perturbación aleatoria.

Para poder aplicar las herramientas estudiadas en la asignatura de econometría necesitamos transformar la expresión de arriba en un modelo lineal. Para ello tomaremos logaritmos en ambos miembros obteniendo:

$$\ln(Y_i) = \ln(A) + \alpha \ln(L_i) + \beta \ln(K_i) + u_i;$$

En definitiva, necesitaremos trabajar no sobre las variables originales sino sobre las variables transformadas a través del logaritmo.

Para el seguimiento de nuestros objetivos de análisis se utilizará R⁵, un entorno de programación y lenguaje para el análisis gráfico y estadístico. El mismo posee herramientas de análisis estadístico (test estadísticos, modelos lineales y no lineales, algoritmos de clasificación y agrupamiento, etc.). Se distribuye bajo las licencias de la Free Software Foundation's GNU General Public License, lo que permite un mayor acceso por parte de profesionales y particulares a un amplio catálogo de herramientas, así como una mayor contribución a aumentar sus librerías, ya que el acceso es gratuito así como las aportaciones técnicas al proyecto R.

Leeremos los datos en R y transformaremos las variables en logaritmos.

Para la determinación de relaciones lineales entre variables haremos uso de gráficos de dispersión proporcionados por R y calculando las correlaciones simples y parciales entre las variables utilizadas.

Para el establecimiento del mejor modelo lineal (regresión lineal múltiple), utilizaremos el estimador máximo verosímil de los parámetros del modelo.

⁵ <http://www.r-project.org/> - The R Project for Statistical Computing

Para analizar la bondad del ajuste haremos uso del error estándar residual, la tabla de ANOVA (Análisis de la Varianza) y el coeficiente de determinación.

Finalmente, las hipótesis del modelo de regresión lineal múltiple (linealidad, normalidad, homocedasticidad, media cero e incorrelación) serán analizadas mediante el test de Shapiro-Wilks, el test de Breusch-Pagan, el test de Durbin-Watson, el histograma de los residuos, el gráfico qq-plot y otra serie de gráficos habituales.

Para todos nuestros análisis asumiremos un nivel de significación no superior al 5%.

Utilizaremos el método Stepwise para analizar un modelo con menos variables, las más representativas.

La guía de referencia utilizada para la elaboración de este análisis estadístico es el manual Modelos Lineales Aplicados en R (Aparicio, Martínez Mayoral, & Morales, Modelos Lineales Aplicados en R), así como los video-tutoriales impartidos en la asignatura de Econometría por el profesor D. Juan Aparicio.

BASE DE DATOS

La base de datos a analizar se encuentra en el siguiente enlace:

http://ec.europa.eu/agriculture/rica/database/database_en.cfm

Esta web da acceso público a la base de datos de la Unión Europea sobre agricultura y desarrollo rural.

La muestra seleccionada para el estudio se describe en la siguiente tabla.

Country	Total labour input (SE010)	Total Utilised Agricultural Area (SE025)	Total livestock units (SE080)	Total specific costs (SE281)	Total farming overheads (SE336)	Rent paid (SE375)	Farm Net Value Added (SE415)	Total fixed assets (SE441)
(BEL) Belgium	2,13	49,14	132,63	119345	45293	10024	91868	573043
(BGR) Bulgaria	2,47	35,71	10,7	14105	9753	5230	18969	70046
(CYP) Cyprus	1,45	9,04	15,89	20343	6993	1019	13361	153327
(CZE) Czech Republic	6,57	227,86	100,78	142816	109530	13369	130048	681928
(DAN) Denmark	1,71	95,26	156,75	209840	81469	18227	167619	2098121
(DEU) Germany	2,22	85,57	89,42	104027	69934	14365	91540	707374
(ELL) Greece	1,15	9,29	6,12	7324	4453	1027	14077	105012
(ESP) Spain	1,4	38,65	22,71	18493	10065	1704	28034	204464
(EST) Estonia	2,01	125,87	38,64	48903	31922	1834	39055	175659
(FRA) France	2,03	85,36	73,17	64664	58196	12828	77253	253675
(HUN) Hungary	1,58	46,26	18,04	29550	17516	3574	31419	104893
(IRE) Ireland	1,22	50,27	59,37	31867	15673	2524	28905	842722
(ITA) Italy	1,26	15,34	12,95	15241	9278	1319	28653	289752
(LTU) Lithuania	1,77	48,49	13,66	17039	8170	1011	18195	77108
(LUX) Luxembourg	1,78	79,09	100,63	73186	53709	9060	60088	947390
(LVA) Latvia	2,02	68,93	20,46	25175	16137	665	19488	87992
(MLT) Malta	1,39	2,62	14,46	21549	6668	230	10289	175260
(NED) Netherlands	2,76	35,65	129,4	181832	113938	12814	150707	1933871
(OST) Austria	1,41	31,46	24,76	20621	23152	2018	31776	346038
(POL) Poland	1,73	18,84	13,02	13261	6098	346	12736	136525
(POR) Portugal	1,59	24,19	14,3	12670	4778	607	15077	83751
(ROU) Romania	1,3	10,05	6,69	4012	2194	553	7084	28621
(SUO) Finland	1,3	54,73	27,98	49016	46695	4124	35285	351539
(SVE) Sweden	1,44	101,27	65,61	99419	61360	12048	53537	743410
(SVK) Slovakia	13,73	521,5	136,44	240305	194596	20482	141208	465473
(SVN) Slovenia	1,46	11,57	12,61	12504	7420	396	5790	179654
(UKI) United Kingdom	2,23	161,13	131,06	128335	54850	10102	87960	1541184

Tabla 3. Base de datos a analizar.

En nuestro caso se han seleccionado una serie de variables y se han descargado los datos en un Excel que luego se ha transformado en un archivo txt para su análisis en R.

La base de datos del año 2012 contiene los siguientes campos:

Country: País (Los datos están referidos a los siguientes países)

- (BEL) Belgium
- (BGR) Bulgaria
- (CYP) Cyprus
- (CZE) Czech Republic
- (DAN) Denmark
- (DEU) Germany

- (ELL) Greece
- (ESP) Spain
- (EST) Estonia
- (FRA) France
- (HUN) Hungary
- (IRE) Ireland
- (ITA) Italy
- (LTU) Lithuania
- (LUX) Luxembourg
- (LVA) Latvia
- (MLT) Malta
- (NED) Netherlands
- (OST) Austria
- (POL) Poland
- (POR) Portugal
- (ROU) Romania
- (SUO) Finland
- (SVE) Sweden
- (SVK) Slovakia
- (SVN) Slovenia
- (UKI) United Kingdom

Total labour input (SE010): Input del trabajo total expresado en AWU (unidad de trabajo anual = equivalente al tiempo completo de un persona). La variable transformada en logaritmo la denominaremos "log(labour_input)".

Total UtilisedAgriculturalArea (SE025): Consiste en la ocupación de la tierra en propiedad, alquilada y compartida (Se expresa en hectáreas 10.000 m²). La variable transformada en logaritmo la denominaremos "log(UtilisedAgriculturalArea)".

Total livestockunits (SE080): Media anual del número de equinos, ganadería, ovejas, cerdos y avícolas en stock. La variable transformada en logaritmo la denominaremos "log(livestockunits)".

Total specificcosts (SE281): Inputs específicos de cosechar (semillas, fertilizantes, etc.), stock de inputs específicos y costes específicos de la silvicultura. La variable transformada en logaritmo la denominaremos "log(specificcosts)".

Total farmingoverheads (SE336): Costes de abastecimiento relacionados con la producción pero no relacionados con líneas específicas de producción. La variable transformada en logaritmo la denominaremos "log(farmingoverheads)".

Rentpaid (SE375): Renta pagada para cultivar, por los edificios y los costes de alquiler. La variable transformada en logaritmo la denominaremos "log(Rentpaid)".

Farm Net ValueAdded (SE415): Remuneración de los factores fijos de producción (trabajo, tierra y capital). Este indicador es sensible a los métodos de producción empleados. La variable transformada en logaritmo la denominaremos "log(Farm_Net_ValueAdded)".

Total fixedassets (SE441): Capital de la tierra cultivable, de los edificios y la silvicultura + Edificios +Maquinaria y equipo + el stock de cría. La variable transformada en logaritmo la denominaremos "log(fixedassets)".

La variable respuesta es FARM_NET_VALUEADDED y el resto son las variables explicativas.



ESTUDIO ECONOMETRICO

Estudio econométrico completo

La variable respuesta es $\log(\text{Farm_Net_ValueAdded})$ y queremos estudiar la relación con el resto de variables mediante la formulación del siguiente modelo:

$$\log(\text{Farm_Net_ValueAdded}) = \beta_0 + \beta_1 \log(\text{labour_input}) + \beta_2 \log(\text{UtilisedAgriculturalArea}) + \beta_3 \log(\text{livestockunits}) + \beta_4 \log(\text{specificcosts}) + \beta_5 \log(\text{farmingoverheads}) + \beta_6 \log(\text{Rentpaid}) + \beta_7 \log(\text{fixedassets}) + u_i$$

Definición de los parámetros:

β_0 : término de interceptación o intercepto.

$\beta_1, \beta_2, \dots, \beta_7$ son los coeficientes de regresión, son las pendientes de sus respectivas variables.

u_i : perturbación aleatoria.

El análisis del modelo se realizará en las siguientes cuatro fases:

1. Análisis preliminar.
2. Ajuste del modelo.
3. Bondad del ajuste.
4. Diagnóstico del modelo.

Siguiendo estos cuatro pasos observaremos la validez del modelo y procederemos a su reajuste.

Una vez realizado el estudio sobre el modelo inicial, propondremos un modelo simplificado con menos variables que volveremos a analizar.

Análisis preliminar

Inicialmente estudiaremos el gráfico de dispersión de la base de datos, con el fin de descubrir de qué tipo son las relaciones entre las variables, si las hay.

Nos dará una visión de la relación lineal entre las variables.

Después de este análisis gráfico realizaremos un análisis numérico de las correlaciones simples y parciales que nos proporcionará datos objetivos de la relación entre variables cuantificando el grado de asociación lineal entre variables.

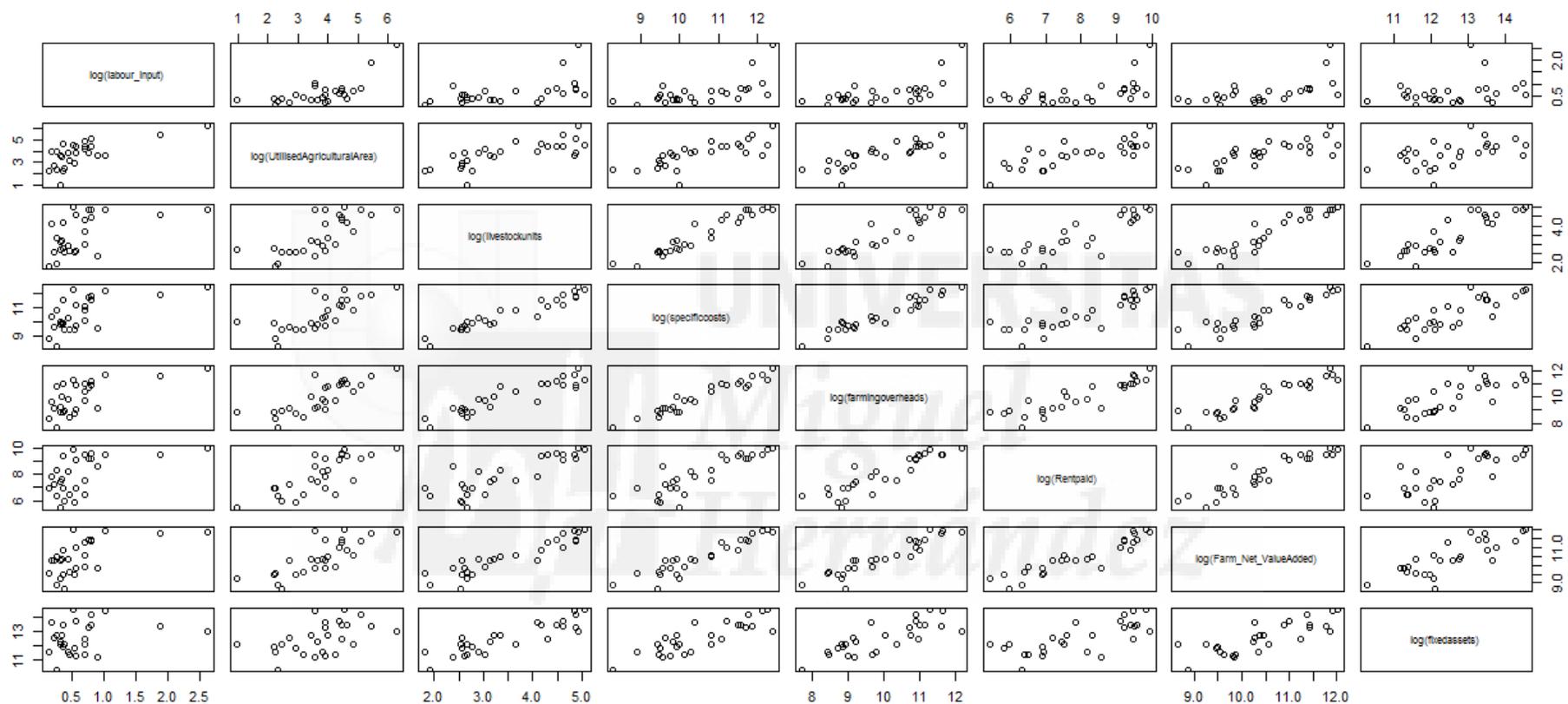


Figura 1. Gráfico de la relación lineal de las variables del modelo.

Podemos observar en la figura 1 en la fila 7 correspondiente a la variable respuesta que las variables explicativas que mejor se ajustan son:

LOG(LIVESTOCKUNITS). Se observa una mayor dispersión en la parte izquierda de la recta.

LOG(SPECIFICCOSTS). Se observa una mayor densidad de puntos en la parte central izquierda de la recta.

LOG(FARMINGOVERHEADS). Se observa una línea creciente más plana en la parte central izquierda de la recta. La recta es la que más linealidad ofrece a la vista.

LOG(RENTPAID). Se observa un hueco en la parte central derecha.

La relación lineal de la variable respuesta y cada una de las variables explicativas se cuantifica calculando la correlación simple de cada una de ellas y están representadas en la siguiente tabla.

	LOG(LABOUR_INP UT)	LOG(UTILISEDAGRICULT URALAREA)	LOG(LIVESTOC KUNITS)	LOG(SPECIFIC COSTS)
LOG(FARM_NET_VAL UEADDED)	0.5779211	0.7830585	0.9227904	0.9353894
	LOG(FARMINGOVE RHEADS)	LOG(RENTPAID)	LOG(FIXEDASS ETS)	
LOG(FARM_NET_VAL UEADDED)	0.9367711	0.9264956	0.8095235	

Tabla 4. Valores de las correlaciones simples.

Tal y como habíamos observado anteriormente en el gráfico, el cálculo de las correlaciones simples nos indica que las variables que más relación tienen con la variable LOG(FARM_NET_VALUEADDED) son:

LOG(LIVESTOCKUNITS) con un 92.28%.

LOG(SPECIFICCOSTS) con un 93.54%.

LOG(FARMINGOVERHEADS) con un 93.68 %.

LOG(RENTPAID) con un 92.65%.

Cuatro de las variables tienen más de un 90% de relación con la variable respuesta y podemos añadir que todas tienen más de un 57%.

A continuación calculamos las correlaciones parciales entre la variable respuesta y cada una de las variables explicativas en presencia del resto de variables.

	LOG(LABOUR_INP UT)	LOG(UTILISEDAGRICULT URALAREA)	LOG(LIVESTOC KUNITS)	LOG(SPECIFIC COSTS)
LOG(FARM_NET_VAL UEADDED)	0.09118	0.08700	0.08899	0.20048
	LOG(FARMINGOVE RHEADS)	LOG(RENTPAID)	LOG(FIXEDASS ETS)	
LOG(FARM_NET_VAL UEADDED)	0.03739	0.57692	0.13347	

Tabla 5. Valores de las correlaciones parciales.

La variable que muestra una mayor relación con LOG(FARM_NET_VALUEADDED) en presencia del resto de variables es LOG(RENTPAID) con un 57.69%.

La información anterior nos ayuda a concluir que pese a que hay varias variables que tienen una gran relación lineal con la variable respuesta, éstas, entre sí, no la tienen tanto dentro del

modelo. Una vez concluyamos el estudio inicial y procedamos a reducir el número de variables del modelo, encontraremos una mejor relación múltiple.

AJUSTE DEL MODELO

En esta segunda fase vamos a calcular los valores que estima el modelo para $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7$, con unos intervalos de confianza al 95% para cada uno de los coeficientes y el contraste de hipótesis asociado a la cuestión de si cada coeficiente es igual o no a cero.

RESUMEN DEL MODELO								
Parámetros	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7
	3.52040	0.06769	0.04072	0.09488	0.25560	0.03756	0.29119	0.08107
p-valor	0.05214	0.69426	0.70769	0.70129	0.38357	0.87216	0.00618	0.56408
Valor residual estándar	0.2853							
Multiple R-squared	0.9359							
Ajusted R-squared	0.9123							
p-value	5.065e-10							

Tabla 6. Valores resumen del modelo.

$$\begin{aligned} \text{LOG(FARM_NET_VALUEADDED)} &= 3.52040 + \text{LOG(LABOUR_INPUT)}*0.06769 + \\ &\text{LOG(UTILISEDAGRICULTURALAREA)}*0.04072 + \text{LOG(LIVESTOCKUNITS)}*0.09488 + \\ &\text{LOG(SPECIFICCOSTS)}*0.25560 + \text{LOG(FARMINGOVERHEADS)}*0.03756 + \\ &\text{LOG(RENTPAID)}*0.29119 + \text{LOG(FIXEDASSETS)}*0.08107 \end{aligned}$$

Los intervalos de confianza al 95% para cada uno de los coeficientes del modelo lineal se muestran en la siguiente tabla.

INTERVALOS DE CONFIANZA		
	Extremo inferior	Extremo superior
Intercepto	-0.03599343	7.0767927
Beta de LOG(LABOUR_INPUT)	-0.28728801	0.4226675
Beta de LOG(UTILISEDAGRICULTURALAREA)	-0.18319452	0.2646402
Beta de LOG(LIVESTOCKUNITS)	-0.41506392	0.6048232
Beta de LOG(SPECIFICCOSTS)	-0.34417413	0.8553764
Beta de LOG(FARMINGOVERHEADS)	-0.44442437	0.5195403
Beta de LOG(RENTPAID)	0.09323051	0.4891524
Beta de LOG(FIXEDASSETS)	-0.20797759	0.3701223

Tabla 7. Valores de los intervalos de confianza.

$$IC_{95\%} \beta_1 = (-0.28728801, 0.4226675)$$

$$IC_{95\%} \beta_2 = (-0.18319452, 0.2646402)$$

$$IC_{95\%} \beta_3 = (-0.41506392, 0.6048232)$$

$$IC_{95\%} \beta_4 = (-0.34417413, 0.8553764)$$

$$IC_{95\%} \beta_5 = (-0.44442437, 0.5195403)$$

$$IC_{95\%} \beta_6 = (0.09323051, 0.4891524)$$

$$IC_{95\%} \beta_7 = (-0.20797759, 0.3701223)$$

Seguidamente estudiaremos el contraste de hipótesis de cada coeficiente para comprobar si pueden ser o no igual a cero.

Los p-valores representados en la tabla Resumen del modelo se analizan a continuación.

Contraste de hipótesis:

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0$$

$$i=1,2,3,\dots,7$$

$$p\text{-valor } \beta_1 = 0,69426 > 0,05$$

Al ser mayor que 0,05 no podemos rechazar H_0 . Esto significa que no es una variable relevante en presencia del resto de covariables.

$$p\text{-valor } \beta_2 = 0,70769 > 0,05$$

Al ser mayor que 0,05 no podemos rechazar H_0 . Esto significa que no es una variable relevante en presencia del resto de covariables.

$$p\text{-valor } \beta_3 = 0,70129 > 0,05$$

Al ser mayor que 0,05 no podemos rechazar H_0 . Esto significa que no es una variable relevante en presencia del resto de covariables.

$$p\text{-valor } \beta_4 = 0,38357 > 0,05$$

Al ser mayor que 0,05 no podemos rechazar H_0 . Esto significa que no es una variable relevante en presencia del resto de covariables.

$$p\text{-valor } \beta_5 = 0,87216 > 0,05$$

Al ser mayor que 0,05 no podemos rechazar H_0 . Esto significa que no es una variable relevante en presencia del resto de covariables.

$$p\text{-valor } \beta_6 = 0,00618 < 0,05$$

Al ser menor que 0,05 se rechaza H_0 . Esto significa que es una variable significativa y asumiremos que el parámetro no es nulo.

$$p\text{-valor } \beta_7 = 0,56408 > 0,05$$

Al ser mayor que 0,05 no podemos rechazar H_0 . Esto significa que no es una variable relevante en presencia del resto de covariables.

Del análisis anterior podemos apreciar que el p-valor relacionado con β_6 0,00618 es inferior a 0,05. La variable relacionada con la β_6 , LOG(RENTPAID), es la variable más importante de la base datos con la mayor correlación parcial y un p-valor menor.

El resto de variables no pueden rechazar la hipótesis H_0 por lo que no son relevantes y no pueden explicar suficientemente la variable respuesta.

BONDAD DEL AJUSTE

Ahora estableceremos si es un buen modelo analizando la bondad del ajuste mediante los tres criterios que se estudian a continuación:

Error estándar residual:

Este criterio se refiere al error cometido a la escala de medida utilizada, por lo que conviene que el error sea lo menor posible.

Una vez estimado el error estándar residual del modelo se determina si es grande o pequeño calculando el coeficiente de variación:

$$cv = (\text{error estándar residual} / \text{la media de LOG(FARM_NET_VALUEADDED)})\%$$

Si el valor del coeficiente de variación es menor al 10% nos indicaría que estamos ante un buen modelo.

En nuestro modelo el error estándar residual es 0.2853.

Calculamos el coeficiente de variación para determinar si estamos ante un buen modelo.

El coeficiente de variación es 2.732154% < 10%. Por lo que en base a este criterio el modelo es bueno.

Tabla de Anova:

Cuantificamos cuánta de la variabilidad contenida en los datos ha conseguido ser explicada por el modelo. Un modelo es bueno cuando la variabilidad explicada es alta y los errores entre los datos y sus predicciones son pequeños.

Realizamos el contraste de hipótesis:

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$$

H_1 : lo contrario

Si el p-valor asociado al contraste de la tabla de ANOVA es inferior a 0,05 rechazaremos H_0 y el modelo será bueno según este criterio.

El p-valor asociado al contraste de la tabla de Anova de nuestro modelo es $5.065e-10$, inferior a 0,05 por lo que rechazamos H_0 y podemos decir que el modelo es bueno según este criterio.

Coefficiente de determinación R^2 :

El coeficiente de determinación es la proporción de la varianza que puede explicarse gracias a la recta de regresión. Mide la capacidad predictiva del modelo ajustado.

Un valor cercano a 1 significará que gran parte de la varianza es explicada a través de la recta de regresión y un valor cercano a cero significará que toda la variabilidad de los datos queda sin explicar por el modelo de regresión.

Por otro lado el coeficiente de determinación ajustado comparado con el coeficiente de determinación nos permitirá elegir el coeficiente de determinación si la diferencia entre ambos es baja y si la diferencia es significativa elegiremos el coeficiente de determinación ajustado.

En nuestro modelo el coeficiente de determinación R_2 es del 93.59% y el ajustado es del 91.23%. Al haber una diferencia pequeña tomaremos el valor R_2 , por lo que el modelo explica en un 93.59% la variabilidad de los datos. Es un valor alto y según este criterio el modelo es bueno.

DIAGNÓSTICO DEL MODELO

Una vez ajustado nuestro modelo, y de haber superado los criterios de bondad del ajuste, verificamos que el modelo satisface las cuatro hipótesis básicas del modelo de regresión lineal para los errores.

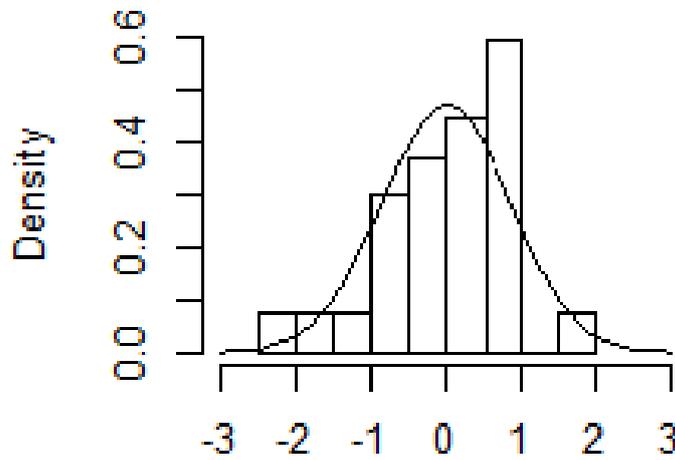
- Normalidad
- Homocedasticidad
- Incorrelación
- Media-cero

El análisis de los residuos nos permitirá estimar si existen o no deficiencias en la comprobación de estas hipótesis, así como descubrir observaciones anómalas o especialmente influyentes en el ajuste.

Se utilizará tanto el análisis gráfico para la fase de diagnóstico, como numérico mediante distintos test que veremos a continuación.

Normalidad

Primeramente comparamos el modelo con la campana de Gauss. Cuanto más se ajuste a la silueta de la campana mayor normalidad tendrá el modelo.



Residuos estandarizados

Figura 2. Histograma.

En el histograma de la figura 2 los valores de la mitad hacia la izquierda guardan cierta relación, pero los valores de la derecha se sobresalen claramente de la silueta de la campana de Gauss. Por lo tanto no podemos decir que se ajusta a una campana de Gauss.

Por otro lado el gráfico qq-plot se representan los residuos ordenados versus los cuantiles correspondientes de una normal. Si la normalidad de los residuos es cierta, los puntos estarán alineados con la diagonal.

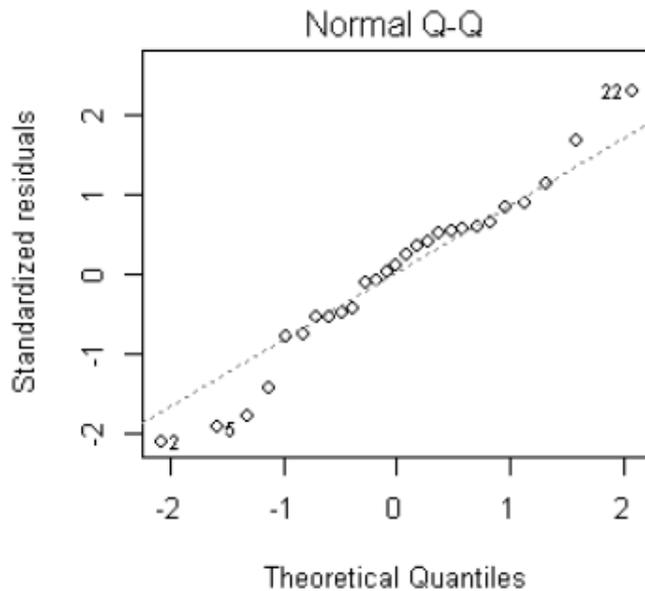


Figura 3. Gráfico QQPLOT.

En la figura 3, vemos que en la parte central de la diagonal los datos están alineados con la diagonal. Pero en la parte de los extremos de la diagonal los puntos están más dispersos. Por lo tanto al igual que ocurría con el histograma, en este gráfico tampoco podemos afirmar que exista normalidad.

El test Shapiro-Wilks nos permite contrastar la normalidad del modelo

H_0 : $e \sim \text{Normal}$

H_1 : lo contrario

El p-valor es de $0.1806 > 0.05$ por lo que no rechazamos H_0 . Esto significa que no rechazamos la hipótesis de Normalidad.

Homocedasticidad

En el caso de que la varianza de los errores sea constante, podremos afirmar que existe homocedasticidad y, en el caso contrario, heterocedasticidad también denominado como problema de varianza no constante, tal problema suele darse cuando el modelo no está bien especificado.

El gráfico que vemos a continuación representa los residuos versus los valores predichos de la variable respuesta.

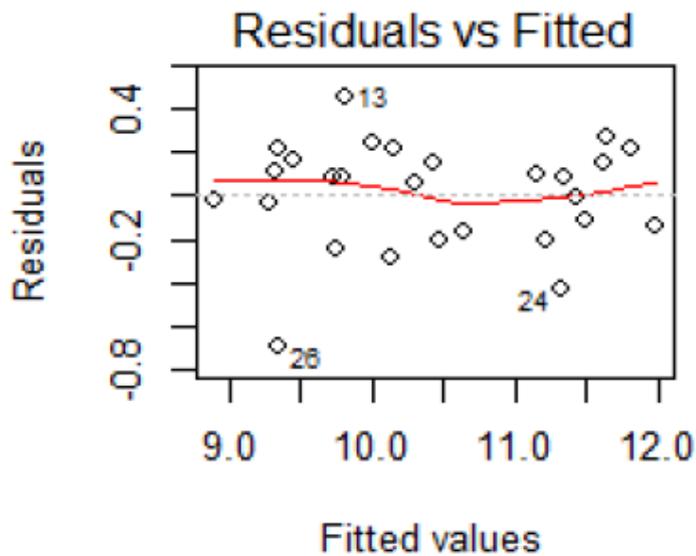


Figura 4. Gráfico de residuos frente a valores predichos.

En la figura 4 lo ideal sería que los residuos aparecieran representados en una banda horizontal sin tendencias alrededor del cero.

El test de Breusch y Pagan nos permite contrastar cuantitativamente la homocedasticidad del modelo:

H_0 : Homocedasticidad

H_1 : Heterocedasticidad

El p-valor es de $0.471 > 0.05$ por lo que no rechazamos H_0 . Esto significa que no rechazamos la hipótesis de Homocedasticidad. Por lo que asumiremos que la varianza es constante y asumiremos que la hipótesis es correcta.

Incorrelación

En este apartado estudiaremos si los valores de unas variables pueden afectar a valores de otras variables, para comprobar esto estudiaremos el gráfico de residuos i frente a residuos $i-1$

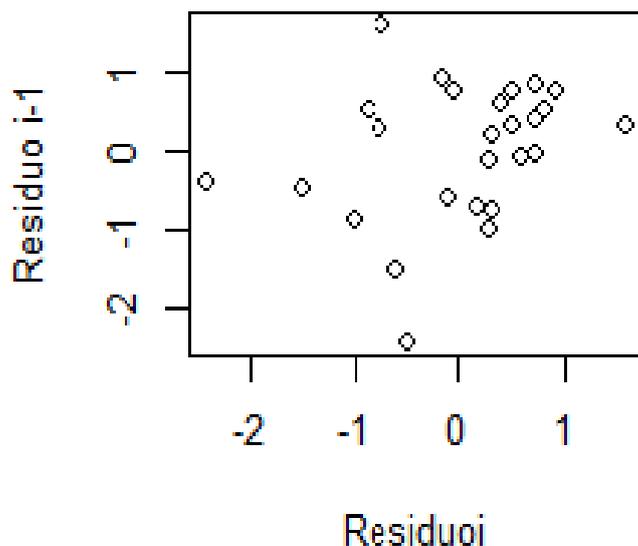


Figura 5. Gráfico de incorrelación.

Se puede observar en la figura 5 una gran incorrelación entre los errores ya que los puntos no guardan una misma distancia.

No aparece ninguna tendencia lineal en la dispersión de los puntos de nuestro modelo.

El test de Durbin y Watson nos permite contrastar la incorrelación de los errores del modelo:

$H_0 : \rho = 0$ (incorrelación)

$H_1 : \rho \neq 0$

El p -valor = $0.1248 > 0.05$ por lo que no rechazamos H_0 . Esto significa que no rechazamos la hipótesis de Incorrelación y, por lo tanto, el modelo es bueno para este criterio.

Media cero

Por último, este criterio siempre se cumple porque utilizamos el modelo de máxima verosimilitud y está implícito que la media de los errores es siempre cero.

Podemos concluir que se cumplen las hipótesis básicas y podemos decir que estamos ante un buen modelo lineal que explica la variable explicativa LOG(FARM_NET_VALUEADDED).

REDUCCIÓN DE VARIABLES

Debido al alto número de variables utilizaremos dos métodos: el Stepwise basado en el *Akaike Information Criteria* y la eliminación de variables de forma manual mediante el criterio del

mayor p-valor. Dichos sistemas nos ayudarán a simplificar el modelo para obtener unos resultados más representativos y que mejor se ajusten al modelo de producción Cobb-Douglas.

El criterio AIC nos indica que valores pequeños son preferidos a valores grandes. Este método nos proporciona un valor AIC inicial, incluyendo todas las variables del modelo, y unos valores AIC en el caso de eliminar alguna de las variables explicativas.

Mediante el Criterio AIC, obtenemos los siguientes valores:

VALOR AIC = -61.22	
VARIABLE	VALOR AIC
LOG(FARMINGOVERHEADS)	-63.178
LOG(UTILISEDAGRICULTURALAREA)	-63.011
LOG(LIVESTOCKUNITS)	-63.001
LOG(LABOUR_INPUT)	-62.990
LOG(FIXEDASSETS)	-62.730
LOG(SPECIFICCOSTS)	-62.108
LOG(RENTPAID)	-52.288

Tabla 8. Stepwise, paso 1.

VALOR AIC = -63.18	
VARIABLE	VALOR AIC
LOG(LIVESTOCKUNITS)	-64.996
LOG(LABOUR_INPUT)	-64.924
LOG(UTILISEDAGRICULTURALAREA)	-64.843
LOG(FIXEDASSETS)	-64.596
LOG(SPECIFICCOSTS)	-62.704
LOG(RENTPAID)	-52.462

Tabla 9. Stepwise, paso 2.

VALOR AIC = -65	
VARIABLE	VALOR AIC
LOG(LABOUR_INPUT)	-66.766
LOG(UTILISEDAGRICULTURALAREA)	-66.438
LOG(FIXEDASSETS)	-65.828
LOG(SPECIFICCOSTS)	-61.862
LOG(RENTPAID)	-54.186

Tabla 10. Stepwise, paso 3.

VALOR AIC = -66.77	
VARIABLE	VALOR AIC
LOG(UTILISEDAGRICULTURALAREA)	-67.945
LOG(FIXEDASSETS)	-67.826
LOG(SPECIFICCOSTS)	-60.942

LOG(RENTPAID)	-56.169
---------------	---------

Tabla 11. Stepwise, paso 4.

VALOR AIC = -67.95	
VARIABLE	VALOR AIC
LOG(FIXEDASSETS)	-69.527
LOG(SPECIFICCOSTS)	-58.387
LOG(RENTPAID)	-53.388

Tabla 12. Stepwise, paso 5.

VALOR AIC = -69.53	
VARIABLE	VALOR AIC
LOG(RENTPAID)	-55.172
LOG(SPECIFICCOSTS)	-51.814

Tabla 13. Stepwise, paso 6.

El criterio AIC ha excluido todas aquellas variables que son menos relevantes y el nuevo modelo resultante es:

$$\text{LOG(FARM_NET_VALUEADDED)} = \beta_0 + \beta_1 \text{LOG(RENTPAID)} + \beta_2 \text{LOG(SPECIFICCOSTS)}$$

Ahora contrastaremos esta reducción de variables con el método de significatividad estadística.

Método de significatividad estadística

Procederemos a la eliminación de covariables mediante el p-valor más alto, esto quiere decir, aquellas covariables en las que su p-valor sea más del 0.05 (tomamos un p-valor del 5% como referencia), serán eliminadas paulatinamente y volveremos a calcular los p-valores resultantes para, paso a paso, ir eliminando covariables hasta encontrarnos en un escenario en el que todas las covariables que dispongan de un p-valor inferior al 5% serán las que formen nuestro modelo final a analizar.

Compararemos las variables resultantes de este método con el Stepwise y *Akaike Information Criteria* para proceder a continuación con el estudio econométrico completo del modelo reducido.

Párametros	p-value
$\beta_0 \log(\text{Farm_Net_ValueAdded})$	0.05214
$\beta_1 \log(\text{labour_input})$	0.69426

$\beta_2 \log(\text{UtilisedAgriculturalArea})$	0.70769
$\beta_3 \log(\text{livestockunits})$	0.70129
$\beta_4 \log(\text{specificcosts})$	0.38357
$\beta_5 \log(\text{farmingoverheads})$	0.87216
$\beta_6 \log(\text{Rentpaid})$	0.00618
$\beta_7 \log(\text{fixedassets})$	0.56408

Tabla 14. P-valor, paso 1.

Eliminamos el p-valor mayor superior a 0.05 correspondiente a la variable $\log(\text{farmingoverheads})$ y volvemos a calcular los p-valores.

Párametros	p-value
$\beta_0 \log(\text{Farm_Net_ValueAdded})$	0.04500
$\beta_1 \log(\text{labour_input})$	0.66842
$\beta_2 \log(\text{UtilisedAgriculturalArea})$	0.62267
$\beta_3 \log(\text{livestockunits})$	0.71670
$\beta_4 \log(\text{specificcosts})$	0.18122
$\beta_6 \log(\text{Rentpaid})$	0.00243
$\beta_7 \log(\text{fixedassets})$	0.51663

Tabla 15. P-valor, paso 2.

Eliminamos el p-valor mayor superior a 0.05 correspondiente a la variable $\log(\text{livestockunits})$ y volvemos a calcular los p-valores.

Párametros	p-value
$\beta_0 \log(\text{Farm_Net_ValueAdded})$	0.000302
$\beta_1 \log(\text{labour_input})$	0.675939
$\beta_2 \log(\text{UtilisedAgriculturalArea})$	0.515295
$\beta_4 \log(\text{specificcosts})$	0.048278
$\beta_6 \log(\text{Rentpaid})$	0.001805
$\beta_7 \log(\text{fixedassets})$	0.346240

Tabla 16. P-valor, paso 3.

Eliminamos el p-valor mayor superior a 0.05 correspondiente a la variable $\log(\text{labour_input})$ y volvemos a calcular los p-valores.

Párametros	p-value
$\beta_0 \log(\text{Farm_Net_ValueAdded})$	0.000233
$\beta_2 \log(\text{UtilisedAgriculturalArea})$	0.418803
$\beta_4 \log(\text{specificcosts})$	0.012524
$\beta_6 \log(\text{Rentpaid})$	0.001530
$\beta_7 \log(\text{fixedassets})$	0.386980

Tabla 17. P-valor, paso 4.

Eliminamos el p-valor mayor superior a 0.05 correspondiente a la variable $\log(\text{UtilisedAgriculturalArea})$ y volvemos a calcular los p-valores.

Párametros	p-value
$\beta_0 \log(\text{Farm_Net_ValueAdded})$	0.000240
$\beta_4 \log(\text{specificcosts})$	0.001902
$\beta_6 \log(\text{Rentpaid})$	0.000202
$\beta_7 \log(\text{fixedassets})$	0.555097

Tabla 18. P-valor, paso 5.

Eliminamos el p-valor mayor superior a 0.05 correspondiente a la variable $\log(\text{fixedassets})$ y volvemos a calcular los p-valores.

Párametros	p-value
$\beta_0 \log(\text{Farm_Net_ValueAdded})$	1.50e-05
$\beta_4 \log(\text{specificcosts})$	3.39e-05
$\beta_6 \log(\text{Rentpaid})$	0.00016

Tabla 19. P-valor, paso 6.

Todos los p-valores son ahora inferiores a 0.05 y podemos comprobar que coinciden con las variables del Stepwise y Akaike Information Criteria.

Obtenemos de esta forma un modelo con exclusivamente dos covariables:

(LOG(RENTPAID)): Renta pagada para cultivar, por los edificios y los costes de alquiler.

(LOG(SPECIFICCOSTS)): Inputs específicos de cosechar (semillas, fertilizantes, etc..), stock de inputs específicos y costes específicos de la silvicultura.

Ahora calculamos la exponencial de la ecuación anterior para convertirla en una función de producción Cobb-Douglas:

$$\text{FARM_NET_VALUEADDED} = \exp\{\beta_0\} * \text{SPECIFICCOSTS}^{\beta_1} * \text{RENTPAID}^{\beta_2}$$

Incorporando los valores de los parámetros tendríamos:

$$\text{FARM_NET_VALUEADDED} = \exp^{2.98095} * \text{SPECIFICCOSTS}^{0.46420} * \text{RENTPAID}^{0.32453}$$

A continuación vamos a realizar el análisis pormenorizado del nuevo modelo simplificado volviendo a transformar las variables en logaritmos para poder hacer el estudio del modelo lineal.

ESTUDIO ECONOMETRICO DEL MODELO SIMPLIFICADO

Al igual que para el estudio completo de regresión múltiple vamos a analizar en las mismas cuatro fases el nuevo modelo simplificado:

- 1.- Análisis preliminar
- 2.- Ajuste del modelo
- 3.- Bondad del ajuste
- 4.- Diagnóstico del modelo

ANÁLISIS PRELIMINAR(Gráfico-Numérico)

El análisis visual del gráfico de dispersión nos muestra una relación lineal de las variables.

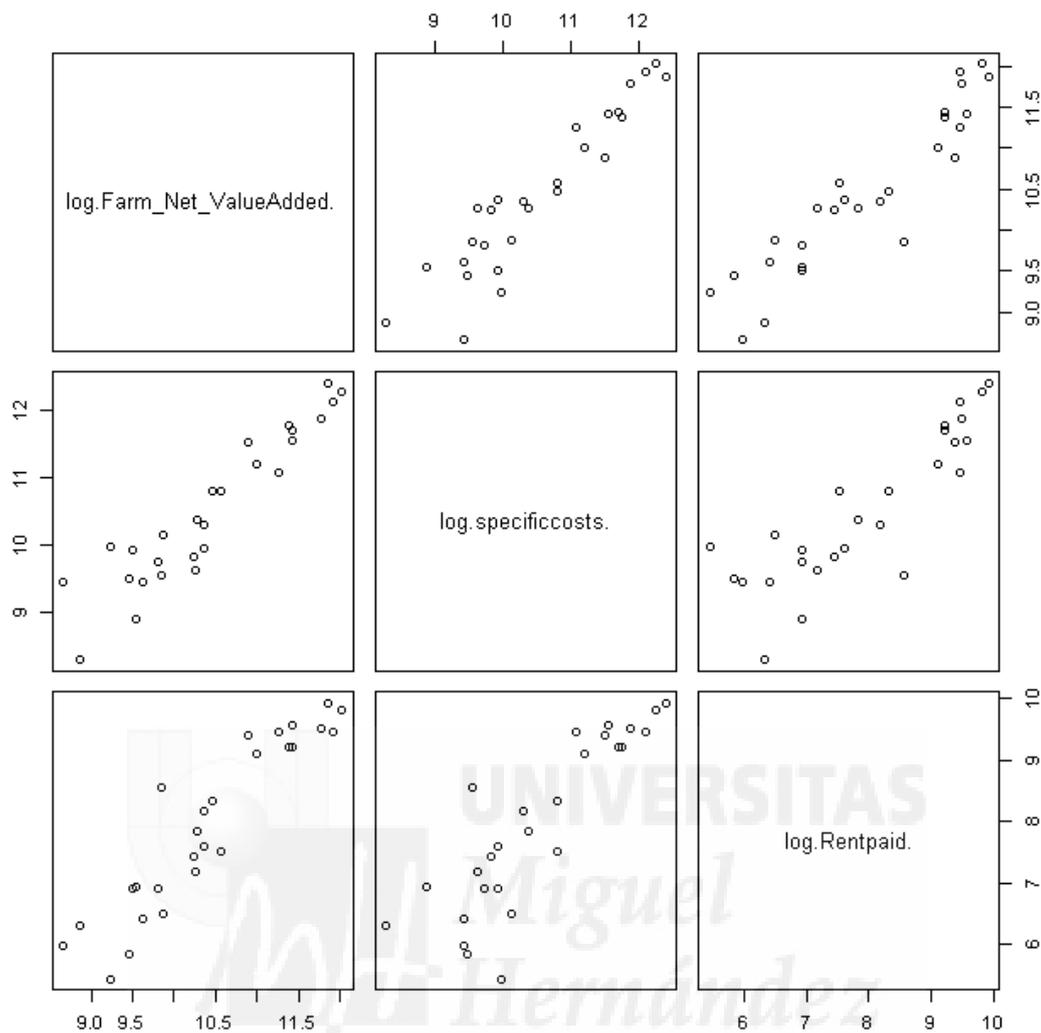


Figura 6. Gráfico de dispersión del modelo simplificado.

En la figura 6 observamos una mayor linealidad con la variable LOG(SPECIFICCOSTS) que con la variable LOG(RENTPAID) cuyos valores centrales están más dispersos.

Las correlaciones se indican en la tabla 19 a continuación:

	LOG(SPECIFICCOSTS)	LOG(RENTPAID)
LOG(FARM_NET_VALUEADDED)	0.9353894	0.9264956

Tabla 20. Valores de las correlaciones simples del modelo simplificado.

Son correlaciones altas que indican una gran linealidad, superiores al 93% y al 92%.

Ahora calculamos las correlaciones parciales que serían:

	LOG(SPECIFICCOSTS)	LOG(RENTPAID)
LOG(FARM_NET_VALUEADDED)	0.71982	0.67404

Tabla 21. Valores de las correlaciones parciales del modelo simplificado.

Las covariables también muestran una gran linealidad en presencia de la otra covariable. Los datos de las correlaciones simples y las parciales son buenas para explicar la linealidad de las variables en el modelo simplificado.

AJUSTE DEL MODELO

Vamos a obtener los valores estimados para β_0 , β_1 y β_2 , así como los intervalos de confianza de los mismos y su contraste de hipótesis. De esta forma identificaremos las variables más importantes de este modelo simplificado.

El modelo económico que deseamos estimar es el siguiente:

$$\text{LOG}(\text{FARM_NET_VALUEADDED}) = \beta_0 + \beta_1 \text{LOG}(\text{SPECIFICCOSTS}) + \beta_2 \text{LOG}(\text{RENTPAID})$$

Estimamos al 95% de confianza

RESUMEN DEL MODELO			
Parámetros	β_0	β_1	β_2
	2.98095	0.46420	0.32453
p-valor	1.50e-05	3.39e-05	0.00016
Valor residual estándar	0.2619		
Multiple R-squared	0.9318		
Ajusted R-squared	0.9261		
p-value	1.019e-14		

Tabla 22. Valores del resumen del modelo simplificado.

Del anterior resumen, podemos describir la estructura del nuevo modelo simplificado a partir de las estimaciones de máximo verosimilitudes de los parámetros, siendo este modelo ajustado el siguiente:

$$\text{LOG}(\text{FARM_NET_VALUEADDED}) = 2.98095 + \text{LOG}(\text{RENTPAID}) * 0.46420 + \text{LOG}(\text{SPECIFICCOSTS}) * 0.32453$$

La siguiente tabla recoge el intervalo de confianza al 95% de cada uno de los parámetros utilizados en el actual modelo:

INTERVALOS DE CONFIANZA		
	Extremo inferior	Extremo superior
Intercepto	1.8422838	4.1196191
Beta de LOG(SPECIFICCOSTS)	0.2756086	0.6527892
Beta de LOG(RENTPAID)	0.1746947	0.4743625

Tabla 23. Valores de los intervalos de confianza del modelo simplificado.

$$IC_{95\%} \beta_0(\text{LOG}(\text{FARM_NET_VALUEADDED})) = (1.8422838, 4.1196191)$$

$$IC_{95\%} \beta_{1(\text{LOG}(\text{SPECIFICCOSTS}))} = (0.2756086, 0.6527892)$$

$$IC_{95\%} \beta_{2(\text{LOG}(\text{RENTPAID}))} = (0.1746947, 0.4743625)$$

Según podemos observar en la Tabla 22, mediante el análisis del intervalo de confianza al 95%, que ninguna de las covariables puede tomar el valor cero.

A continuación realizamos el contraste de hipótesis:

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0$$

$$i=1,2,3,\dots,7$$

p-valor $\beta_1 = 3.39e-05$. Al ser menor que 0.05 se rechaza H_0 . Esto significa que es una variable significativa y asumiremos que el parámetro no es nulo.

p-valor $\beta_2 = 0.00016$. Al ser menor que 0.05 se rechaza H_0 . Esto significa que es una variable significativa y asumiremos que el parámetro no es nulo.

Del análisis anterior podemos apreciar que el p-valor de ambas variables es inferior a 0.05, por lo que ambas son importantes para este modelo simplificado.

BONDAD DEL AJUSTE

Ahora estableceremos si es un buen modelo analizando la bondad del ajuste.

Error estándar residual:

En el modelo el error estándar residual es 0.2619 . Tendremos que calcular el coeficiente de variación para determinar si es bueno.

El coeficiente de variación es $2.508066 < 10\%$. Por lo que en base a este criterio el modelo es bueno.

Tabla de Anova:

Cuantificamos cuánta de la variabilidad contenida en los datos ha conseguido ser explicada por el modelo.

$$H_0: \beta_1 = \beta_2 = 0$$

$$H_1: \text{lo contrario}$$

$$p\text{-value} = 1.019e-14$$

El valor es inferior al 0.05 por lo que rechazamos H_0 y podemos decir que el modelo es bueno para este criterio.

Coeficiente de determinación R²:

El coeficiente de determinación R² es del 93.18% y el ajustado es del 92.61%. Por lo que hay poca diferencia entre ambos y los valores son muy altos y explican la mayoría de los datos del modelo.

Cuanto más cercano a 1 (100%) sea el valor, más parte de la varianza es explicada por la recta de regresión.

DIAGNÓSTICO DEL MODELO

Una vez ajustado nuestro modelo, y de haber superado los criterios de bondad del ajuste, verificamos que el modelo satisface las cuatro hipótesis básicas del modelo de regresión lineal para los errores.

- Normalidad
- Homocedasticidad
- Incorrelación
- Media-cero

El análisis de los residuos nos permitirá estimar si existen o no deficiencias en la comprobación de estas hipótesis, así como descubrir observaciones anómalas o especialmente influyentes en el ajuste.

Se utilizará tanto el análisis gráfico para la fase de diagnóstico, como numérico mediante distintos test que veremos a continuación.

En la bondad del ajuste se cumplen todos los criterios. Ahora verificamos que el modelo satisface las hipótesis básicas.

Normalidad

En el gráfico qq-plot se representan los residuos ordenados versus los cuantiles correspondientes de una normal. Si la normalidad de los residuos es cierta, los puntos estarán alineados con la diagonal.

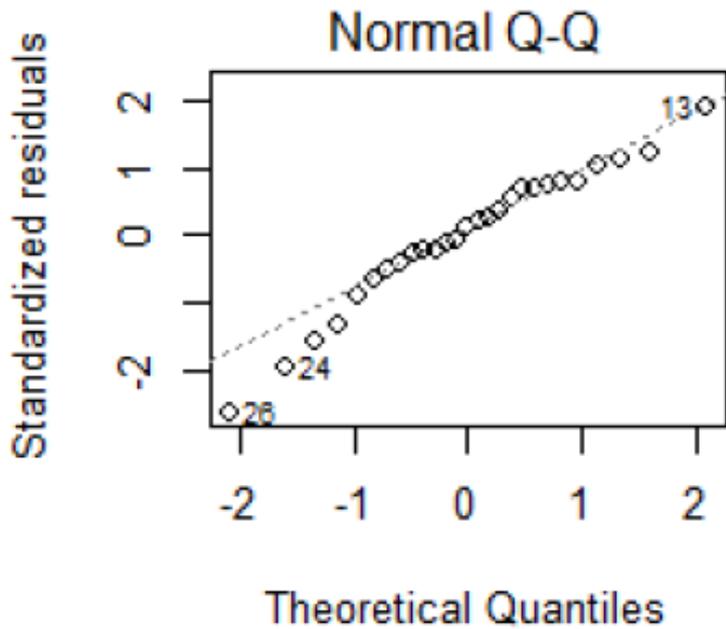


Figura 7. Gráfico QQPLOT del modelo simplificado.

En los valores centrales se observa una relación lineal muy aproximada con la normalidad, los valores iniciales muestran una gran dispersión y los valores finales un algo de dispersión.

Ahora comparamos el modelo simplificado con la campana de Gauss. Cuanto más se ajuste a la silueta de la campana mayor normalidad tendrá el modelo.

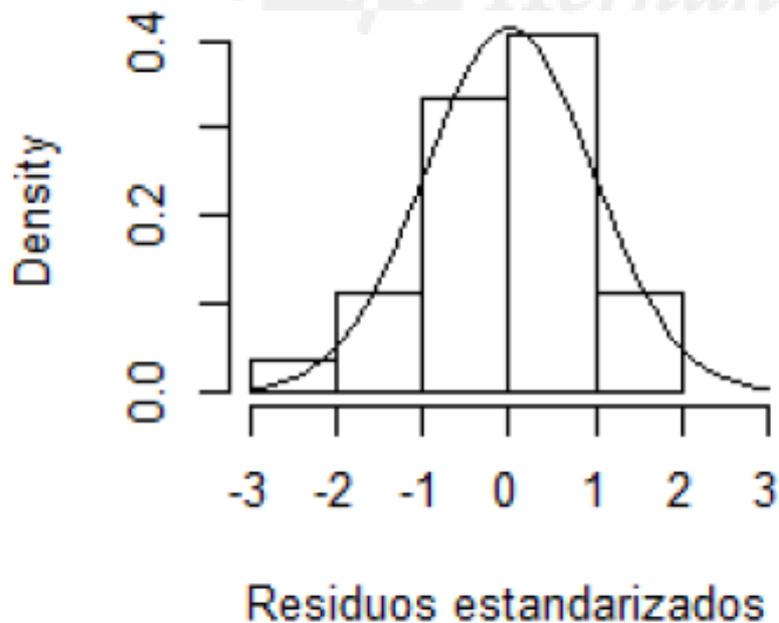


Figura 8. Histograma del modelo simplificado.

No guarda relación con una campana de Gauss salvo en la mitad izquierda.

El test Shapiro-Wilks nos permite contrastar la normalidad del modelo

H_0 : $e \sim \text{Normal}$
 H_1 : lo contrario

El p-valor es de $0.482 > 0.05$ por lo que no rechazamos H_0 . Esto significa que no rechazamos la hipótesis de Normalidad.

Homocedasticidad

En el caso de que la varianza de los errores sea constante, podremos afirmar que existe homocedasticidad y, en el caso contrario, heterocedasticidad también denominado como problema de varianza no constante, tal problema suele darse cuando el modelo no está bien especificado.

La figura 9 el gráfico que vemos a continuación representa los residuos versus los valores predichos de la variable respuesta en el modelo simplificado.

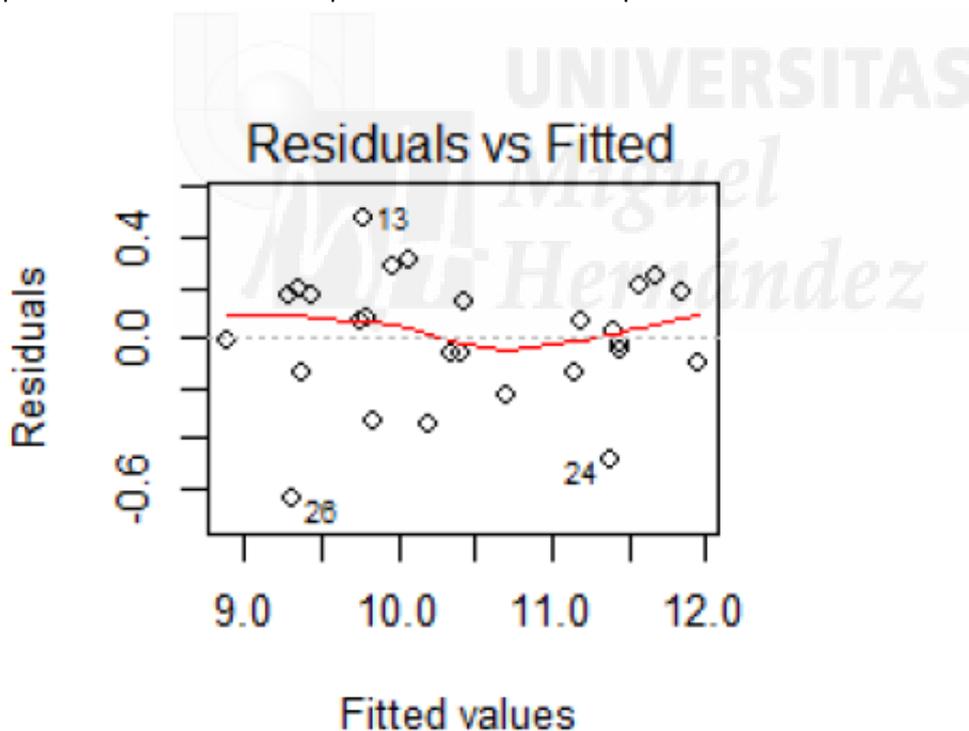


Figura 9. Gráfico de residuos frente a valores predichos del modelo simplificado.

Podemos observar que la variabilidad del error es alta los puntos están muy dispersos.

El test de Breusch y Pagan nos permite contrastar la homocedasticidad del modelo.

H_0 : Homocedasticidad
 H_1 : Heterocedasticidad

El p-valor es de $0.5317 > 0.05$ por lo que no rechazamos H_0 . Esto significa que no rechazamos la hipótesis de Homocedasticidad. Según este criterio el modelo es bueno.

Incorrelación

Estudiaremos a continuación si los valores de unas variables pueden afectar a valores de otras variables, para comprobar esto observaremos el gráfico de residuos i frente a residuos $i-1$

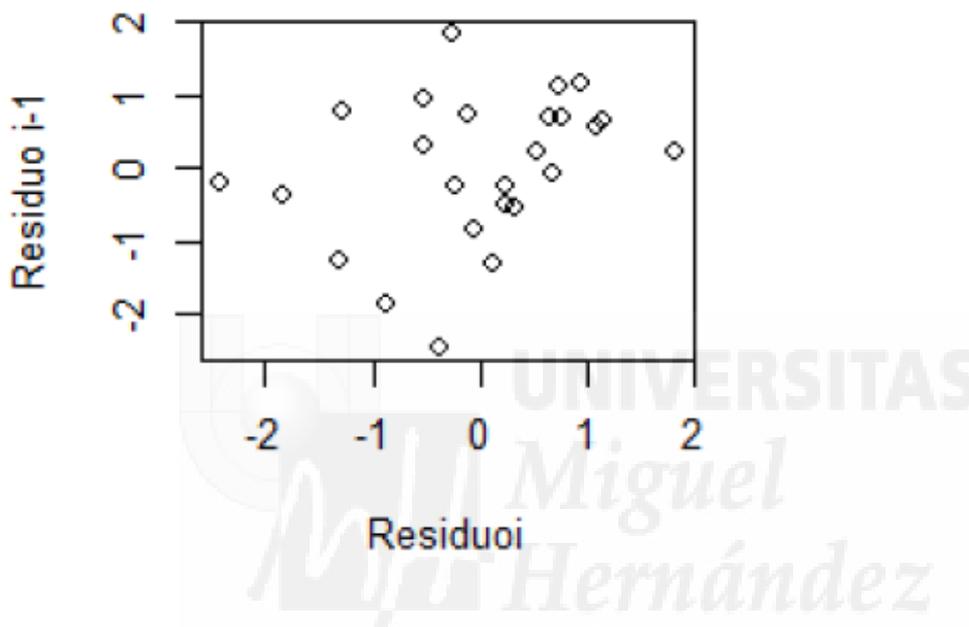


Figura 10. Gráfico de incorrelación del modelo simplificado.

En la figura 10 se puede observar una incorrelación entre los errores ya que los puntos no guardan una misma distancia y no guardan ninguna relación.

El test de Durbin y Watson nos permite contrastar la incorrelación de los errores del modelo.

$H_0 : \rho = 0$ (incorrelación)

$H_1 : \rho \neq 0$

El p-valor = $0.09135 > 0.05$ por lo que no rechazamos H_0 . Esto significa que no rechazamos la hipótesis de Incorrelación. Según este criterio el modelo es bueno.

Media cero

Este criterio siempre se cumple porque utilizamos el modelo de máxima verosimilitud y está implícito que la media de los errores es siempre cero.

Por lo tanto se cumplen las hipótesis básicas y podemos decir que estamos ante un buen modelo lineal.



CONCLUSIONES

Para finalizar este trabajo describo las conclusiones a continuación.

El objetivo es estimar la función de producción de una base de datos de agricultura europea y para ello hemos transformado las variables en logaritmos para hacer un estudio del modelo lineal econométrico y poder obtener las variables que se relacionan mejor con la variable respuesta Farm Net ValueAdded, a través de una función de producción tipo Cobb-Douglas.

El estudio econométrico del modelo completo ha sido positivo ya que el resultado de los distintos criterios así lo han sido y nos hace concluir que el modelo completo es bueno.

A continuación mediante los métodos de reducción de variables Stepwise-Akaike Information Criteria y de significatividad estadística hemos reducido el modelo original a uno que considera exclusivamente dos covariables RENTPAID y SPECIFICCOSTS. Este modelo simplificado lo hemos analizado y la conclusión también ha sido positiva. Según los distintos criterios econométricos el modelo es realmente bueno para explicar la variable respuesta.

El modelo simplificado con las variables transformadas en logaritmos quedaría como sigue:

$$\text{LOG}(\text{FARM_NET_VALUEADDED}) = \beta_0 + \beta_1 \text{LOG}(\text{SPECIFICCOSTS}) + \beta_2 \text{LOG}(\text{RENTPAID})$$

$$\text{LOG}(\text{FARM_NET_VALUEADDED}) = 2.98095 + 0.46420 \text{LOG}(\text{SPECIFICCOSTS}) + 0.32453 \text{LOG}(\text{RENTPAID})$$

Para representar el modelo de producción calculamos la exponencial del modelo simplificado. Quedando de la siguiente manera:

$$\text{FARM_NET_VALUEADDED} = e^{2.98095} + \text{SPECIFICCOSTS}^{0.46420} + \text{RENTPAID}^{0.32453}$$

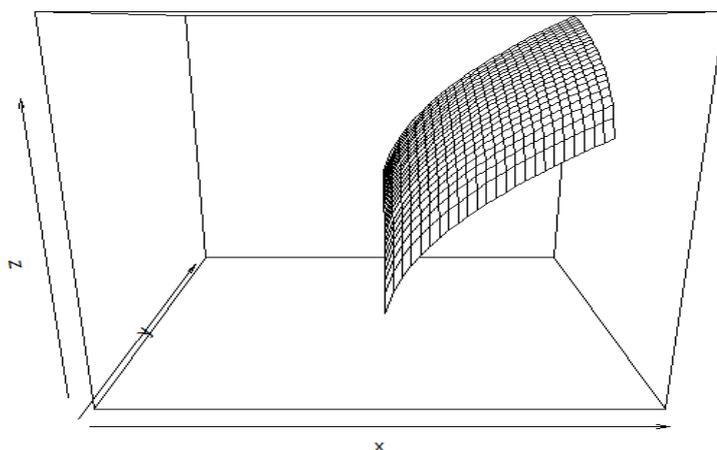


Figura 11. Gráfica de la función de producción de la agricultura europea (Fuente: elaboración propia).

La elasticidad Total de Producción (ϵ) sería:

$$\beta_1 + \beta_2 = 0.78873$$

Por lo tanto, al tomar valor menor que uno los rendimientos a escala detectados serían decrecientes. Esto quiere decir que el incremento que se produce en FARM_NET_VALUEADDED al modificar SPECIFICCOSTS y/o RENTPAID es menor que el incremento de ambas. Esto de alguna manera nos indica que existe cierto grado de saturación en la relación del uso de inputs y del output. Por lo que la obtención de mejores tasas de rendimiento por unidad de insumo invertida debe venir de la mano de cambios tecnológicos y no tanto de incrementos sustanciales en los inputs. Existen así diseconomías de escala en la producción respecto a los factores considerados, al constatar la existencia de rendimientos decrecientes de escala. Algo similar fue detectado en un entorno agrario de nivel nacional (DÍAZ PALOMARES, MARTÍNEZ PAZ y VICARDO MODROÑO 2003).

En conclusión, el valor añadido neto del sector agrario en la Unión Europea puede explicarse de forma adecuada a través de dos variables: por un lado, los insumos específicos de cosechar (semillas, fertilizantes, etc..) junto con el stock de inputs específicos y costes específicos de la silvicultura, y, por otro lado, la renta pagada por cultivar (edificios y costes de alquiler).



REFERENCIAS

APARICIO, J.; MARTÍNEZ, M.; MORALES, J. *Modelos Lineales Aplicados en R*. Universidad Miguel Hernández. 2004.

Comisión Europea. Dirección General de comunicación. [en línea]. <http://europa.eu/pol/index_es.htm> [Consulta: 15 de septiembre de 2015]

DÍAZ PALOMARES, R., J.M. MARTÍNEZ PAZ, y V. VICARDO MODROÑO. «Eficiencia versus innovación en explotaciones agrarias.» *Estudios de Economía Aplicada*, 2003: 485-501.

The R Project for Statistical Computing. Licencia GNU. [en línea]. <<http://www.r-project.org/>> [Consulta 5 de marzo de 2015]

