

Trabajo Fin de Grado

Emisiones de dióxido de carbono a nivel mundial: un estudio estadístico de su relación con las principales magnitudes macroeconómicas



Autor: Patricia Climent Sánchez

Tutor: Juan Aparicio Baeza

Fecha de entrega: septiembre 2018

Curso Adaptación G.A.D.E.

Facultad CC Sociales y Jurídicas de Orihuela

Universidad Miguel Hernández

ÍNDICE

1.- Resumen.	5
2.- Introducción	6
2.1.- Crecimiento vs. Ingresos: La curva de Kuznets.	8
2.2.- La curva de Kuznets: el caso español.	9
2.3.- Desarrollo sostenible y las autoridades ambientales.	11
3.- Objetivos.....	12
4.- Metodología.....	13
4.1.- Muestra de datos	14
4.2.- Definición de las variables	16
5.- Estudio econométrico	16
5.1.-Análisis preliminar.....	17
5.2.-Ajuste del modelo	20
5.3.-Bondad del ajuste	23
a)Error estándar residual	23
b)Tabla de Anova.....	24
c)Coeficiente de determinación o R^2	25
5.4.-Diagnóstico del modelo.....	25
a)Normalidad	26
b)Homocedasticidad	28
c)Incorrelación	30
d)La media de los errores es cero.....	31
6.- Transformación del modelo	31
a)Normalidad	33
b)Homocedasticidad	35
c)Incorrelación	36
d)La media de los errores es cero.....	38
7.- Procedimiento secuencial de selección de variables	38
8.- Estimación de la respuesta media de cada país.....	39
9.- Conclusiones	41
10.- Bibliografía	45

ÍNDICE DE GRÁFICOS Y TABLAS

Gráficos:

Gráfico 1: Curva de Kuznets Ambiental	8
Gráfico 2: Diagrama o gráfico de dispersión	18
Gráfico 3: Gráfico QQ-plot del modelo completo	27
Gráfico 4: Histograma de residuos del modelo completo	27
Gráfico 5: Gráfico de residuos vs. valores ajustados del modelo completo.....	29
Gráfico 6: Gráfico de incorrelación del modelo completo	30
Gráfico 7: Gráfico Q-Q Plot del modelo transformado.....	33
Gráfico 8: Histograma de residuos del modelo transformado	34
Gráfico 9: Gráfico de residuos estandarizados vs. valores ajustados del modelo transformado.....	35
Gráfico 10: Gráfico de incorrelación del modelo transformado	37

Tablas:

Tabla 1: Muestra utilizada para el estudio estadístico	15
Tabla 2: Correlaciones simples del modelo completo	19
Tabla 3: Correlaciones parciales del modelo completo	20
Tabla 4: Ajuste del modelo completo	21
Tabla 5: Intervalos de confianza del modelo completo	22
Tabla 6: Test de normalidad Shapiro-Wilks del modelo completo	28
Tabla 7: Test de homocedasticidad de Breusch-Pagan del modelo completo	29
Tabla 8: Test de incorrelación de Durbin-Watson del modelo completo	31
Tabla 9: Test de normalidad Shapiro-Wilks del modelo transformado	34
Tabla 10: Test de homocedasticidad de Breusch-Pagan del modelo Transformado	36
Tabla 11: Test de incorrelación de Durbin-Watson modelo transformado ...	37
Tabla 12: Tabla del criterio AIC	38
Tabla 13: Intervalos de confianza de la estimación de la respuesta media.....	40

1.- Resumen.

El objeto principal de este análisis es el de estudiarse si existe una correlación entre ciertas variables que miden o cuantifican la contaminación en el aire y diversas magnitudes macroeconómicas básicas que tienen que ver con el nivel de desarrollo o riqueza de un país o territorio, tales como el stock de capital, el número de trabajadores (en miles) y el valor añadido.

Con el propósito de poder acceder a toda la información disponible, se toma como base los precios del año 1995, que -si bien- no es una fecha muy actual, nos permite extraer conclusiones sobre el modelo presentado, y poder aplicarlas más adelante en futuros estudios.

Para ello, voy a utilizar un modelo de regresión lineal múltiple, en el que aplicaré cuatro fases, que son:

- (i) Análisis preliminar.
- (ii) Ajuste del modelo.
- (iii) Bondad del ajuste.
- (iii) Diagnóstico del modelo.

Como consecuencia de los problemas de normalidad que presenta el modelo, procederé a aplicar una transformación Box-Cox, con el objeto de estudiar si es posible mejorar los resultados del modelo completo, para finalizar con el procedimiento secuencial de selección de variables.

Las principales conclusiones extraídas del modelo expuesto y desarrollados son que las tres variables estudiadas -*stock de capital*, el *número de trabajadores* (en miles) y el *valor añadido* (tomando como base precios del año 1995) tienen relación y son relevantes para el modelo, por lo que no deben ser eliminadas, de forma que la variabilidad de la variable respuesta, emisiones de CO₂, puede ser explicada a través de estas tres covariables.

2.- Introducción

La contaminación es un problema global de la sociedad en la que vivimos, y -como tal- está sometido a continuos estudios y medidas por parte de los gobernantes de los países desarrollados. El propio desarrollo económico, y la prosperidad asociada a él, parece que esté unido de forma inseparable al concepto de contaminación. De hecho, prácticamente cualquier actividad que llevemos a cabo en nuestro quehacer diario, contamina. Por ejemplo, al encender el aire acondicionado, la calefacción, utilizar el coche o cualquier otro medio de transporte, las actividades económicas como la producción agrícola e industrial, son acciones que producen emisiones de CO₂ en la atmósfera, y nadie se plantea a estas alturas que puedan ser eliminadas o sustituidas de nuestra vida cotidiana. Resultaría impensable que se prohibiera conectar el aire acondicionado en verano, o prohibir la posesión de un vehículo privado. Las consecuencias serían desastrosas, y no ya tanto en el ámbito de la comodidad del consumidor, sino en el ámbito económico.

Como primera premisa al respecto, parece claro que puede afirmarse, sin ninguna reserva, que este tipo de contaminación está relacionada con el crecimiento económico y la población: con un mayor ingreso per cápita se consume más, lo cual, en condiciones normales y sin la presencia de otras variables que perturben el resultado, va a producir más residuos. De hecho, el llamado *problema ambiental* ya ha sido objeto de numerosos estudios, ensayos y comentarios, existiendo dos principales corrientes, de pensamiento opuesto, ante los problemas ambientales: la corriente *pesimista* cree que, si se mantiene la tendencia actual se provocará un problema ecológico-económico. Por el contrario, la corriente *optimista*, cree que el propio crecimiento económico corrige los problemas ambientales de forma automática, al terminar creciendo de forma más eficiente.

Es conocido por todos los continuos aumentos en los niveles de contaminación atmosférica, produciendo fenómenos como la lluvia ácida, el calentamiento global y el empeoramiento de la capa de ozono. Aprovechando el ejemplo de la capa de ozono, también es cierto que las medidas tomadas por los gobernantes al respecto (tales como eliminación de

sprays que contengan ciertos gases) ha hecho que el eventual *agujero* haya disminuido notablemente.

Tratando de clasificar la contaminación en subtipos, y según estudios llevados a cabo, pueden distinguirse dos tipos de contaminación: *primaria* y *secundaria*.

- Contaminación primaria: es un contaminante que se emite directamente al aire.

- Contaminación secundaria: es un contaminante que no es emitido directamente, sino que se forma en la atmósfera cuando reaccionan otros contaminantes (contaminantes primarios).

La contaminación preocupa a todos los países (al menos en teoría) y los gobiernos, generalmente, han actuado en materia medioambiental. Se busca lograr un desarrollo sostenible de forma que se modifiquen y adapten nuevos hábitos de consumo y de vida, consiguiendo una concienciación en el individuo.

Son numerosos los libros que hacen referencia a qué actividades rutinarias que llevamos a cabo día a día podemos modificar para ayudar a reducir el problema de la contaminación, como es el caso de *100 gestos para salvar el mundo*¹ y *Eco ¡Lógico!: ¡Únete al debate medioambiental!*², lecturas que debemos llevar a cabo para concienciarnos aún más si cabe. Desde un punto de vista más especializado, podemos encontrar otro libro de interesante lectura: *Contaminación ambiental. Una perspectiva desde la química*³, donde se habla de la gestión ambiental y se realiza un estudio de los contaminantes.

¹Johan Tell. 100 gestos para salvar el mundo. ART BOOKS EDICIONES SAS (2013)

² Joanna Yarrow. Eco ¡Lógico!: ¡Únete al debate medioambiental! Editorial Blume. (2010)

³José Marcos Alfayate y otros. Contaminación ambiental. Una perspectiva desde la química. Editorial Rústica .(2008)

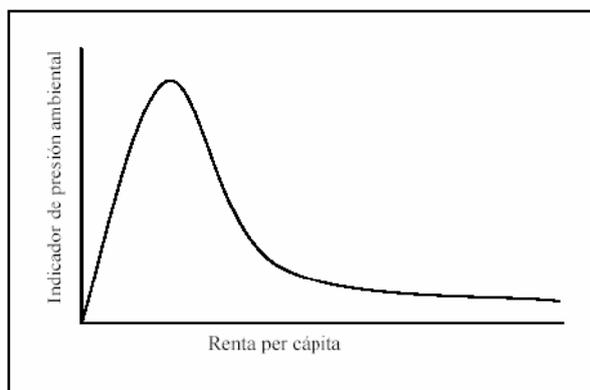
2.1.- Crecimiento vs. Ingresos: La curva de Kuznets.

Según el artículo publicado en la Revista Desarrollo Local Sostenible⁴, Simón Kuznets (1901-1985), Premio Nobel de Economía en 1971, formuló una teoría para relacionar el crecimiento económico y los ingresos per cápita. Según esta teoría, el propio crecimiento económico reduciría la desigualdad en los ingresos. Como se realizarían inversiones por parte de las economías, generarían empleo y aumentaría la productividad, lo que daría como resultado salarios más altos y una redistribución del ingreso per cápita.

Esta hipótesis de la curva de Kuznets fue aplicada al campo ambiental a partir de estudios llevados a cabo, en primer lugar, por Shafrik y Bandyopadhyay (1992) y el Banco Mundial (1992). En estos estudios se indicaba que ciertas emisiones de contaminantes seguían el comportamiento gráfico similar al de una U invertida.

En 1993, Panayotou, en sus estudios, introduce la hipótesis de la Curva de Kuznets Ambiental (en adelante CKA) donde explica que la contaminación aumenta con el crecimiento económico hasta cierto nivel de ingreso, después del cual empieza a reducirse (Arrow et al., 1995; Suri y Chapman 1998; Ekins, 2000).

Gráfico 1: Curva de Kuznets Ambiental



Fuente: Revista de Desarrollo Local Sostenible

⁴Revista Desarrollo Local Sostenible. Grupo Eumed.net y Red Académica Iberoamericana Local Global. Vol 3, Nº 8. Recuperado el 27 de mayo de 2018 de <http://www.eumed.net/rev/delos/08/>

Las conclusiones a las que llega la llamada Curva de Kuznets Ambiental (CKA) son que, a partir de un cierto nivel de desarrollo económico, es el propio crecimiento económico el que se encarga de solucionar los problemas ambientales, puesto que los propios países van a demandar mejoras ambientales que les permitan continuar creciendo de forma sostenible y sostenida en el tiempo. Del mismo modo, relaciona el concepto de libre comercio como elemento que genera, de forma indirecta, mejoras ambientales, puesto que considera que el libre comercio es el marco óptimo para un mayor y mejor crecimiento económico y que esta mejora desembocará finalmente en una mejora ambiental provocada por la propia necesidad de continuar creciendo de forma sostenible.

2.2.- La curva de Kuznets: el caso español.

Para el seguimiento de este apartado, bucearemos en el trabajo expuesto en la Revista de Economía Industrial⁵ en la que en un estudio publicado para el caso español por Esteve y Tamarit (2012a, 2012b) se presenta la relación entre las variables PIB per cápita y emisiones de CO₂. En dicho estudio, se indica que la relación según la curva de Kuznets ambiental (CKA) se encuentra en una fase de degradación ambiental, aunque se observa un cambio de comportamiento, lo que nos llevaría a pensar que sí que se cumple la CKA para el caso de España.

Existen estudios para el período 1850-2000 como los de Rubio (2005), Carpintero (2005) o Roca y Alcántara (2001) que revelaban el rechazo a dicha hipótesis debido a que se obtuvo como resultado que las emisiones de CO₂ estaban determinadas por el crecimiento económico y, sin embargo, no se produce el cambio en la relación entre producción y degradación ambiental que postula la CKA.

También hay que añadir que estos estudios comentados se han llevado a cabo desde el lado de la oferta mientras que el trabajo realizado por Ramos–

⁵Revista de Economía Industrial N.º 389. Recuperado el 3 de mayo de 2018 de <http://www.minetad.gob.es/Publicaciones/Publicacionesperiodicas/EconomiaIndustrial/RevistaEconomiaIndustrial/389/Jes%C3%BAAs%20Iglesias.pdf>

Martín (2003) incorpora el estudio desde el lado de la demanda (como por ejemplo el consumo de hogares) llegando a la conclusión de que España no sigue la hipótesis formulada en la CKA.

Otro estudio a tener en cuenta es el de Díaz-Vázquez y Cancelo (2010) donde de nuevo se confirma que, para España, la evidencia de la CKA no se cumple.

Como puede observarse, no hay unanimidad al respecto para el caso español, pero mi conclusión personal es que -de los distintos estudios examinados- es cierto que en el más reciente sí que encuentra una correlación como la descrita en la CKA, por lo que es probable que, para el caso español, sí se cumpla la CKA; lo que sucede es que hace relativamente pocos años que hemos empezado la fase decreciente de la curva, esto es, que hasta hace poco, nuestra economía no estaba suficiente desarrollada, globalizada y liberalizada para que se notaran los efectos descritos por Kuznets, y que ha sido a raíz de ciertas medidas, tales como la adopción del Euro como moneda, el espacio único europeo, y otras medidas económicas de liberalización y privatización (fin de los monopolios energéticos, de transportes, de suministros, etc.) cuando hemos empezado a comportarnos como una economía avanzada en la que internamente se exige mayor control ambiental sin renunciar al crecimiento en ningún caso.

Como reflexión de un tema que está en boca de todos en los últimos meses, podemos citar el tema de la contaminación en ciudades como Madrid y Barcelona, las cuales han adoptado medidas para fomentar el uso de vehículos más eficientes y menos contaminantes, y prohibiendo la circulación de los vehículos más antiguos y -por tanto- menos eficientes. Estas medidas están provocando la paulatina sustitución de los vehículos diésel, por otros híbridos, eléctricos, etc. Con ello, se va a generar más consumo (renta) y a la vez, menos contaminación.

2.3.- Desarrollo sostenible y las autoridades ambientales.

A final de los años ochenta se comenzó a dar a conocer un nuevo término que todos los países han adaptado y asumido: desarrollo sostenible.

Este concepto se puede definir según el manual del Módulo de Sensibilización Ambiental⁶ como “aquel modelo de desarrollo que busca satisfacer las necesidades del presente sin comprometer la satisfacción de las necesidades de generaciones futuras”.

Para ello, se necesita llevar a cabo cambios en los sistemas de producción de las industrias, productos, mercados e incluso en el comportamiento del ser humano cambiando su forma de vida y costumbres.

Con el objetivo de alcanzar el desarrollo sostenible, todos los países se han puesto en marcha y han creado una red de autoridades para la consecución de dicho objetivo.

En España, por ejemplo, se ha creado un organismo denominado *Red de Autoridades Ambientales*. Se trata de un foro de cooperación que coordina las actuaciones medioambientales con respecto a la gestión y destino de los fondos comunitarios: se trata de aplicar y velar por el cumplimiento de las medidas de protección de medioambiente en la ejecución de programas financiados con fondos europeos.

Está formada por los representantes de las Comunidades Autónomas que se designen como las Consejerías de Medio Ambiente y las Consejerías de Hacienda, el Ministerio de Medio Ambiente que coordina y hace de secretario, la Federación Española de Municipios y Provincias que representa a la Administración Local y por la Administración General del Estado que:

-A través del Ministerio de Hacienda controla los fondos FEDER y Fondo de Cohesión.

-A través del Ministerio de Agricultura, Pesca y Alimentación, los Fondos

⁶ VV.AA. Manual del Módulo de Sensibilización Ambiental. Edita Ministerio de Trabajo y Asuntos Sociales y Ministerio de Medio Ambiente. ISBN:84-8320-217-4. Recuperado el 08 de mayo de 2018 de <http://www.mapama.gob.es/es/calidad-y-evaluacion-ambiental/temas/red-de-autoridades-ambientales-raa-/sensibilizacion-medioambiental/>

FEOGA e IFOP.

-A través del Ministerio de Trabajo y Asuntos Sociales, el fondo FSE.

A nivel europeo, existe un organismo similar al español en cuanto a objetivos, denominado Red Europea de Autoridades Ambientales y de Autoridades de Gestión (ENEA-MA), el cual aglutina las distintas redes de autoridades de los Estados miembros de la UE, encontrándose entre sus objetivos el del intercambio de información entre los países, constituyendo grupos de trabajo sobre las materias que se consideren más importantes, y velar por el cumplimiento de los requisitos ambientales en todos los sectores manteniendo un mínimo de dos reuniones anuales.

De forma parecida a la composición de España, pero a nivel europeo, está formado por la Comisión Europea a través de la Dirección General de Medio Ambiente y las Direcciones Generales que controlan y administran los Fondos comunitarios como:

- La Dirección General de Política Regional controla los fondos FEDER y Fondo de Cohesión.
- La Dirección General de Agricultura, el fondo FEOGA.
- La Dirección General de Empleo y Asuntos Sociales, el fondo FSE.
- La Dirección General de Pesca, el fondo IFOP.

3.- Objetivos.

El principal objetivo del presente estudio es comprobar de un modo contrastado, si la llamada variable respuesta, que -en este caso- son las emisiones de CO₂, está influenciada por ciertas variables macroeconómicas como el stock de capital, el número de trabajadores y el valor añadido.

Para ello va a utilizarse un modelo de regresión lineal múltiple, en el que se efectuará el análisis a variables macroeconómicas de distinta nacionalidad, así como, estudiar el efecto que tienen en nuestro modelo y determinar si son importantes estadísticamente para explicar la variable respuesta.

4.- Metodología.

En el estudio que vamos a llevar a cabo utilizaremos herramientas gráficas y numéricas asumiendo, en cada uno de los métodos, un margen de error no superior al 5%.

Para analizar las relaciones entre la variable y las covariables de nuestro modelo, procederemos a estudiar, en primer lugar, los gráficos de dispersión y las correlaciones simples y parciales y confirmaremos la importancia de esas covariables en el modelo según el criterio AIC (Akaike InformationCriteria). El llamado Criterio AIC es una medida que nos ayuda a seleccionar el mejor modelo basándose en la bondad del ajuste y la relación existente entre el sesgo y la varianza. Se tomará como mejor modelo el único conjunto de parámetros que maximice la función de verosimilitud (menor dato AIC).

La bondad del ajuste la estudiaremos a través del error estándar residual, el análisis de la varianza (la tabla de ANOVA) y el coeficiente de determinación.

El último paso que realizaremos será comprobar si se cumplen las cuatro hipótesis básicas: normalidad, homocedasticidad, incorrelación y la media de los errores es cero. Para esto utilizaremos el test de Shapiro –Wilk, el test de Breusch-Pagan, el test de Durbin- Watson, el histograma de residuos y el gráfico QQ-plot.

Todo este estudio estadístico se ha llevado a cabo utilizando el software libre The R Project for Statistical Computing⁷ distribuido bajo la licencia GNU GPL. Es un programa que proporciona una amplia gama de herramientas estadísticas, actualizado constantemente debido a que sus propios usuarios pueden desarrollar nuevos paquetes o mejoras, e incorporarlas al programa a modo de actualizaciones. Esta característica, junto a su capacidad gráfica de alta calidad, contribuye a que sea uno de los lenguajes más utilizados en la investigación.

Junto a este software, se ha utilizado como guía, el manual Modelos

⁷The R Project for Statistical Computing. <http://www.r-project.org>

Lineales Aplicados en R (Aparicio, Martínez, & Morales, 2004)⁸ y los videos disponibles en el material de la asignatura Econometría de la página web de la Universidad Miguel Hernández del profesor D. Juan Aparicio Baeza⁹.

4.1.- Muestra de datos

Para llevar a cabo este estudio se ha utilizado un registro de treinta y ocho países, algunos de ellos pertenecientes a la OCDE, actualizados sus datos en el año 2005. Para ello tendremos una variable respuesta, emisiones de CO₂ y tres covariables o variables explicativas, el stock de capital, el número de trabajadores y el valor añadido tomando como base precios del año 1995.

Los datos han sido obtenidos de la base de datos WIOD (*World Input-Output Database*), publicada por Timmer y otros en 2015, como resultado de un proyecto financiado por la Unión Europea, el cual contiene series anuales de datos de cuarenta países, del periodo 1995 a 2011, de variables de producción y medioambientales¹⁰.

La base de datos estudiada se presenta en la siguiente tabla:

⁸Aparicio, J., Martínez, M., & Morales, J. (2004). Modelos Lineales Aplicados en R. Universidad Miguel Hernández

⁹<http://umh3067.edu.umh.es/material>

¹⁰Aparicio J. y otros. Journal of Environmental Management 196 (2017) páginas 148 – 160.

Tabla 1: Muestra utilizada para el estudio estadístico.

Country	Emissions_of_CO2	Capital_stock	Number_of_employees	Gross_value	OECD
Australia	402300,243	1839240,462	8740,9381	487723,7032	YES
Austria	74683,1	890264,4278	3352,668	223442,7768	YES
Belgium	131355,91	928896,6322	3563,414	256238,3839	YES
Bulgaria	52077,22	7510,1595	2522,7	1624,4791	NO
Brazil	310158,7792	2968849,409	72221,4912	735855,012	NO
Canada	552329,0359	2754075,271	14846,975	866365,7825	YES
Cyprus	7978,54	47289,0984	285,8	12659,0639	NO
Czech_Republic	118297,47	600706,5608	4084,331	118907,5002	YES
Germany	898557,2	9036087,657	34480	2207153,942	YES
Denmark	85473,39	570265,4453	2553,816	123834,7122	YES
Spain	364888,73	3061284,074	16458	761480,4773	YES
Estonia	16686,07	42460,4223	558,6	9705,1446	YES
Finland	62272,92	415381,457	2118,8	122402,7969	YES
France	427363,18	6347803,628	22874,8	1435304,44	YES
United_Kingdom	621585,19	3841033,763	27093	1387682,618	YES
Greece	113168,23	528896,536	2918,401	163142,967	YES
Hungary	61846,92	273583,2305	3627,318	54662,7107	YES
Indonesia	350116,1861	746726,8723	36447,4136	327480,9941	NO
India	1197919,873	4745935,309	199766,7741	1948212,785	NO
Ireland	47512,12	296013,6981	1629,775	93839,1086	YES
Italy	499357,6	4684524,001	18359,4	1124487,439	YES
Japan	1206901,409	15260657,35	53676,0551	4270387,733	YES
Korea_Republic_of	533906,3358	2713827,083	15186	795719,8134	YES
Lithuania	16095,91	112052,3804	1210,7	29315,8356	NO
Luembourg	5356,5	76343,9069	287,8	22627,0939	YES
Latvia	8564,58	46258,8238	909,4122	14558,1444	NO
Mexico	413758,5446	1113095,218	27857,6081	387838,8105	YES
Malta	2608,25	21024,8295	133,1179	6055,9864	NO
Netherlands	206417,06	1573116,74	7105,241	397937,3557	YES
Poland	324045,2	790190,57	10436,3	234888,7313	YES
Portugal	71780,4	502019,976	4128,1	137954,4921	YES
Romania	106363,27	23329,9543	6162,2	6261,8756	NO
Russia	1635392,026	308629,6488	63551,3388	150049,0371	NO
Slovak_Republic	41370,45	176866,3395	1818,74	45317,2254	YES
Slovenia	17406,35	63657,256	760,349	21407,7367	YES
Sweden	62685,76	565868,5058	4103,2	229495,4708	YES
Turkey	248898,0898	40903,2134	12473,3415	19990,6856	YES
United_States	5529034,166	27863993,32	139145,6428	10484168,07	YES

Fuente: Elaboración propia

4.2.- Definición de las variables

Las variables utilizadas en este estudio son:

-Emisiones de dióxido de carbono (Emissions_of_CO2): son los vertidos, causados por el hombre, que se realizan en la atmósfera de dicho gas incoloro, denso y poco reactivo (CO₂).

-Stock de capital (Capital_stock): es una variable fundamental para el estudio del crecimiento económico. Es un conjunto de bienes que forman parte del proceso productivo de bienes o servicios.

-Número de trabajadores (Number_of_employees): es el número de personas que, con el fin de obtener una recompensa (sueldo), presta servicios a una empresa.

-Valor añadido tomando como base precios del año 1995 (Gross_value): es el incremento de valor que se genera en un bien por causa de su transformación durante el proceso productivo.

5.- Estudio econométrico

El modelo teórico objeto de ajuste, es un modelo de regresión lineal múltiple, donde la variable respuesta (y) es emisiones de CO₂, y las variables explicativas (X) o covariables, el stock de capital, el número de trabajadores y el valor añadido tomando como base de precios el año 1995.

La formulación de nuestro modelo es:

$$\text{Emissions_of_CO2} = \beta_0 + \beta_1 * \text{Capital_stock} + \beta_2 * \text{Number_of_employees} + \beta_3 \text{Gross_value} + \varepsilon$$

Definición de parámetros:

β_0 : Intercepto.

β_1, β_2 y β_3 son los coeficientes de regresión, esto es, las pendientes de sus respectivas variables.

ε : es el error que se comete en cada observación cuando hacemos uso del modelo.

El análisis del modelo se va a realizar a través de cuatro fases:

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

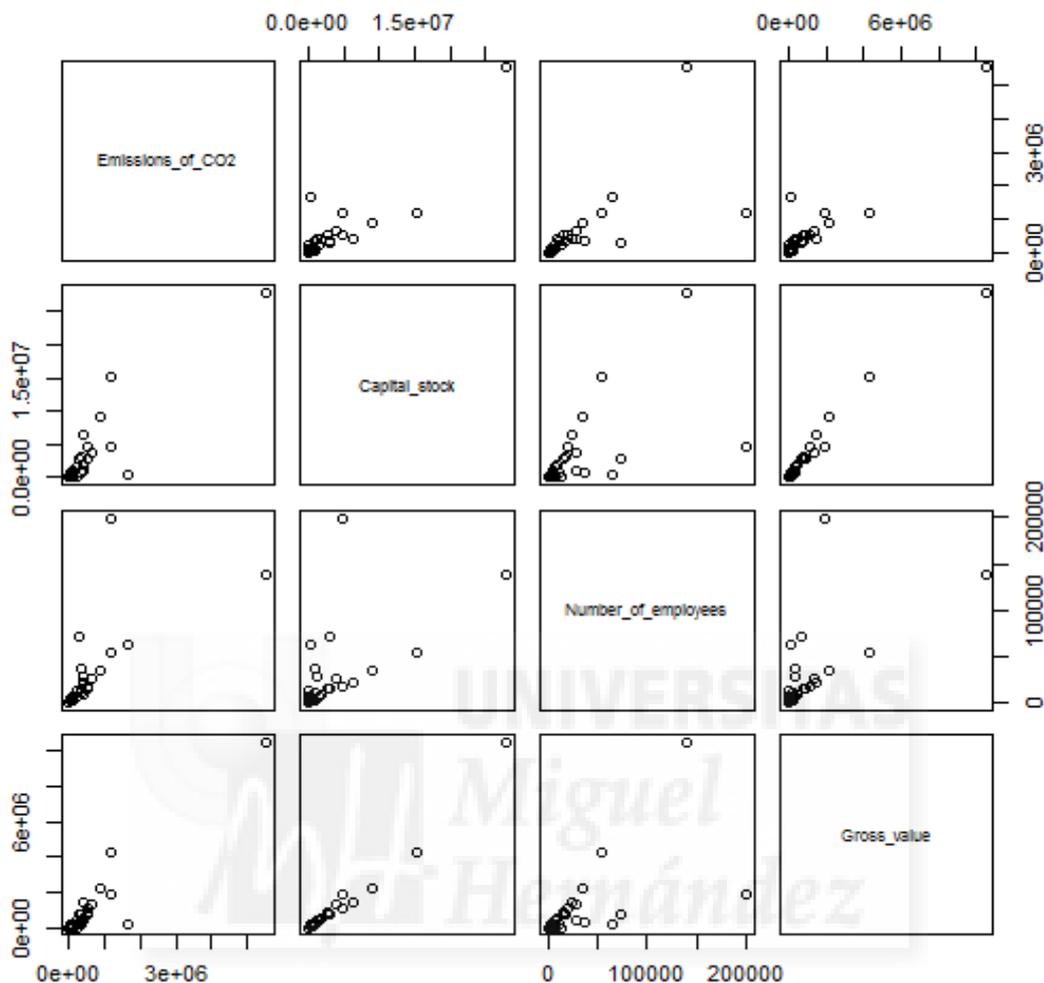
5.1.-Análisis preliminar

En primer lugar, realizamos el análisis preliminar que se divide en dos partes, una gráfica y una numérica.

Comenzamos con la parte gráfica estudiando y analizando el gráfico de dispersión, de forma que, podremos obtener una idea inicial de la relación existente entre la covariable o variable respuesta (Y) y las variables explicativas (X), para continuar con la parte numérica donde, a través de las correlaciones parciales y simples comprobaremos si se cumple el resultado obtenido gráficamente.

Los resultados obtenidos en la correlación simple pueden no ser correctos o estar incompletos respecto a la relación de las covariables y la variable respuesta, para lo cual, realizamos la ampliación del estudio a las correlaciones parciales.

Gráfico 2: Diagrama o gráfico de dispersión.



Fuente: elaboración propia.

En el gráfico 2 hemos obtenido los datos de dispersión de las covariables y su relación con la variable respuesta. Esto es, la relación entre las variables Capital_stock, Number_of_employees y Gross_value con respecto a Emissions_of_CO2.

Las tres covariables presentan una nube de puntos con tendencia creciente y lineal respecto de la variable emisiones de CO₂, aunque, en todos los casos, se encuentran outliers que nos pueden sugerir dudas respecto a la existencia de que dicha relación entre las variables sea una relación lineal.

Para comprobar si se cumple esta linealidad, vamos a estudiar las correlaciones simples y parciales a través del análisis numérico.

Las correlaciones simples son:

Tabla 2: Correlaciones simples del modelo completo

	Emissions_of_CO2	Capital_stock	Number_of_employees	Gross_value
Emissions_of_CO2	1.0000000	0.8951692	0.6980003	0.9404866

Fuente: Elaboración propia

Estudiando estas correlaciones simples de la tabla 2, podemos decir, que existe unacorrelación muy alta, 0.9404866 entre las emisiones de CO₂ y el valor añadido tomando como base precios del año 1995, al igual que ocurre con el stock de capital, 0.8951692.

Respecto al número de trabajadores (en miles) también existe esta correlación, aunque no es tan elevada como en los casos anteriores, 0.6980003.

Con los datos que hemos obtenido, podemos confirmar la relación creciente encontrada de forma gráfica.

Obtenemos ahora las correlaciones parciales:

Tabla 3: Correlaciones parciales del modelo completo.

	Emissions_of_CO2	Capital_stock	Number_of_employees	Gross_value
Emissions_of_CO2	0.0000000	-0.53316	0.30829	0.74431

Fuente: Elaboración propia

Estudiando las correlaciones parciales expuestas en la tabla 3 podemos decir que existe unacorrelación parcial alta, 0.74431, entre las emisiones de

CO₂ y el valor añadido tomando como base precios del año 1995, lo que significa que esta covariable explica mucho de nuestra variable respuesta emisiones de CO₂ una vez descontado todo lo que explican las otras variables.

Con respecto al stock de capital, -0.53316, podemos observar que esta covariable también explica bastante de nuestra variable respuesta, aunque en menor medida.

Respecto al número de empleados, la correlación es baja por lo que no aporta a nuestro modelo información adicional y la relación lineal es poco evidente.

Con el análisis preliminar podemos concluir que es muy probable que podamos establecer un buen modelo lineal en los siguientes apartados.

5.2.-Ajuste del modelo

Para realizar el ajuste del modelo vamos a seguir tres pasos. Comenzaremos calculando los valores para las tres covariables y sus intervalos de confianza al 95% y a partir de ahí, trabajaremos con los datos obtenidos de sus p-valor para resolver el contraste de hipótesis respecto a si el coeficiente asociado a dicha variable puede ser nulo, lo que indicaría que desaparecería de nuestro modelo espacio. Los resultados obtenidos son:

Tabla 4: Ajuste del modelo completo.

Parámetros		P-valor
β_0	98570000,00	0.066593
β_1	-183,40	0.000814
β_2	2678,00	0.067350
β_3	957,40	1.95e-07
Otros valores:		
Valor residual estandar	259900 on 34 degrees of freedom	
Multiple R-squared	0.928	
Adjusted R-squared	0.9216	
p-value	< 2.2e-16	

Fuente: Elaboración propia

De los datos de la tabla 4 obtenemos que nuestro modelo es:

$$\text{Emissions_of_CO2} = 98570000 - 183.40 * \text{Capital_stock} + 2678 * \text{Number_of_employees} + 957.40 * \text{Gross_value} + \epsilon$$

Esto significa que, si mantengo igual el resto de las variables X y varío en 1 unidad esa covariable, Emissions_of_CO2 aumenta o disminuye en el valor del β . Esto es, por ejemplo, que si aumentamos en 1 unidad Capital_stock, Emissions_of_CO2 disminuye en 183.40 unidades y si, por el contrario, aumentamos en 1 unidad Number_of_employees o Gross_value, Emissions_of_CO2 aumenta en 2678 unidades para el caso del número de trabajadores y en 957.40 unidades para el valor añadido tomando como base precios del año 1995.

Tabla 5: Intervalos de confianza del modelo completo.

	2,50%	97,50%
Intercepto	-7127.2357065	204265.6
Capital_stock	-0.2848652	-0.08198357
Number_of_employees	-0.2020186	5.558004
Gross_value	0.6580326	1.256848

Fuente: Elaboración propia

De los datos de la tabla 5 obtenemos que los intervalos de confianza para el modelo completo son:

$$IC95\%(\beta_0) = (-7127.2357065, 204265.6)$$

$$IC95\%(\beta_1) = (-0.2848652, -0.08198357)$$

$$IC95\%(\beta_2) = (-0.2020186, 5.558004)$$

$$IC95\%(\beta_3) = (0.6580326, 1.256848)$$

Estos valores son los datos entre los que se mueve el valor de nuestra variable respuesta al aumentar en 1 unidad una variable X manteniendo el resto constante. Esto es, por ejemplo, que si aumento en 1 unidad Gross_valuemanteniendo el resto de las variables constantes, Emissions_of_CO2 aumentará un valor entre 0.6580326 y 1.256848, siendo el valor de dicho β_3 la media del intervalo. Pero si, aumentamos en una unidad Capital_stockmanteniendo el resto de las variables constantes, Emissions_of_CO2 disminuirá un valor entre 0.2848652 y 0.08198357, siendo el valor de dicho β_1 la media del intervalo. En el caso que aumentemos una unidad en la variable Number_of_employees podemos obtener como resultado que Emissions_of_CO2 aumente o disminuya puesto que se mueve en un intervalo de -0.2020186 a 5.558004.

Una vez analizados estos datos, continuamos realizando el contraste de hipótesis para comprobar si cada uno de los coeficientes β puede ser igual o no a cero. Como β_0 no acompaña a ninguna variable, no puede hacer que desaparezca nuestro modelo por lo que analizamos solo los coeficientes que acompañan a variables.

Para ello, trabajamos con los datos de los p-valor obtenidos con anterioridad y expuestos en la tabla 4.

El ajuste del modelo haciendo el contraste de hipótesis es:

$$\left. \begin{array}{l} H_0: \beta_1 = 0 \\ H_1: \beta_1 \neq 0 \end{array} \right\} \text{ p-value: } = 0.000814 < 0.05$$

Este valor nos indica que hay evidencias en los datos para rechazar H_0 a favor de H_1 . Por lo que la variable Capital_stock es necesaria para el modelo y no podemos eliminarla, en presencia del resto de covariables.

$$\left. \begin{array}{l} H_0: \beta_2 = 0 \\ H_1: \beta_2 \neq 0 \end{array} \right\} \text{p-value:} = 0.067350 > 0.05$$

Este valor nos indica que no hay suficientes evidencias en los datos como para rechazar H_0 a favor de H_1 . Por lo que la variable Number_of_employees no es significativa en el modelo en presencia del resto de covariables.

$$\left. \begin{array}{l} H_0: \beta_3 = 0 \\ H_1: \beta_3 \neq 0 \end{array} \right\} \text{p-value:} = 1.95e-07 < 0.05$$

Este valor nos indica que hay evidencias en los datos para rechazar H_0 a favor de H_1 . Por lo que la variable Gross_value es necesaria para el modelo y no podemos eliminarla, en presencia del resto de covariables.

Como conclusión podemos decir que no es significativa para nuestro modelo la variable Number_of_employees, lo que significa que, cuando están las otras variables en el modelo, esta no añade información nueva y significativa sobre la variable respuesta (contaminación).

5.3.-Bondad del ajuste

Para evaluar lo bueno que es nuestro modelo y si puede predecir el valor de la variable respuesta realizamos la bondad del ajuste con sus tres criterios.

a) Error estándar residual

Es una medida para estudiar la bondad del ajuste referente a la escala de medida utilizada. Preferimos modelos donde el error estándar residual

estimado sea pequeño, para lo cual utilizamos el coeficiente de variación, esto es, inferior al 10%. Si el coeficiente de variación es menor al 10%, indica que estamos ante un buen modelo.

Para nuestro caso, siguiendo los datos de la tabla 4 vemos que el error estándar residual es 259900. Una vez calculamos el coeficiente de variación obtenemos que su valor es 58.7%. Este valor lo consideramos elevado, está por encima del 10% recomendado por el Instituto Nacional de Estadística, por lo que debemos decir que nuestro modelo no aprueba el primer criterio de la bondad del ajuste.

b) Tabla de Anova

Con este análisis queremos comprobar si nuestro modelo es capaz de explicar la variabilidad contenida en nuestros datos. Si la variabilidad es alta y las diferencias entre los datos y las predicciones son pequeñas podremos decir que estamos ante un buen modelo.

Para calcularlo contestamos al contraste de hipótesis de forma que nuestro p-value debe ser inferior a 0.05 lo que indicaría que rechazaríamos H_0 a favor de H_1 y nuestro modelo sería bueno según este criterio.

Para nuestro caso, el contraste de hipótesis sería:

$$\left. \begin{array}{l} H_0: \beta_1 = \beta_2 = \beta_3 = 0 \\ H_1: \text{Lo contrario} \end{array} \right\} \text{p-value: } < 2.2e-16 < 0.05$$

Siguiendo los datos de la tabla 4 vemos que el p-valor asociado al criterio de la tabla Anova es menor a $2.2e-16$. Este valor nos indica que hay evidencias en los datos para rechazar H_0 a favor de H_1 por lo que debemos decir que nuestro modelo aprueba el criterio de la tabla de Anova.

c) Coeficiente de determinación o R^2

El coeficiente de determinación o R^2 es la proporción de la varianza que es explicada por la recta de regresión. Obtenemos un valor que nos indicará, en tanto por cien, la variabilidad total de los datos explicada por el modelo de regresión. Si el valor es cercano a uno indicará que buena parte de la varianza es explicada por la recta de regresión y, por el contrario, cuanto más cercano a cero significará que prácticamente toda la variabilidad de los datos queda sin explicar por el modelo de regresión.

Además, debemos comparar los datos de R^2 con los de R^2 ajustado y ver si son parecidos o no.

Siguiendo los datos de la tabla 4 vemos que el valor de R^2 es 0.928, esto significa que nuestro modelo de regresión consigue explicar el 92.8 % de la variabilidad total de los datos. Es un valor muy alto por lo que podemos decir que es un buen modelo.

Comprobamos ahora R^2 ajustado para ver si las variables de nuestro modelo son correctas y no hay introducidas variables de forma artificial. De los datos de la tabla 4 obtenemos que R^2 ajustado es de 0.9216. Dado que los valores son muy parecidos podemos analizar cualquiera de los dos con el mismo resultado, por lo que podemos decir que es un buen modelo.

Como resumen de la bondad del ajuste debemos decir que nuestro modelo no aprueba el criterio del error estándar residual, aunque sí que aprueba y con buena nota, los criterios de la tabla Anova y del coeficiente de determinación.

5.4.-Diagnóstico del modelo

Tras estudiar la bondad del ajuste de nuestro modelo con las deficiencias detectadas en cuanto a no cumplir el criterio del error estándar residual, vamos a comprobar y verificar, si en este caso, se satisfacen las hipótesis básicas del modelo de regresión lineal, que son;

- a) Normalidad
- b) Homocedasticidad
- c) Incorrelación
- d) La media de los errores es cero

A lo largo de este análisis podremos comprobar si existen deficiencias en estos cuatro criterios presentados de forma que, en el caso de que hubiese alguna, lleváramos a cabo, el replanteamiento del modelo y sus variables.

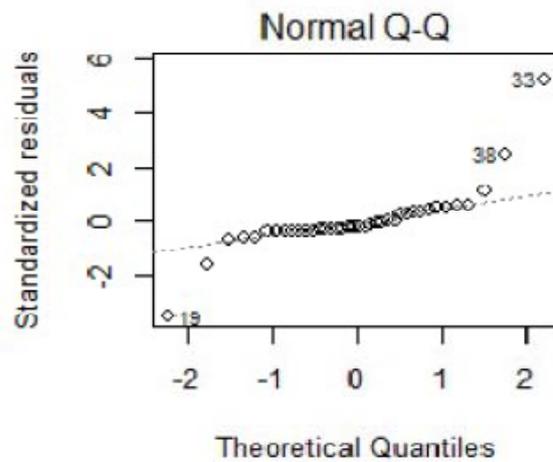
A partir de aquí procedemos al diagnóstico del modelo para lo cual vamos a utilizar, en todos sus criterios, una parte gráfica y otra parte numérica de datos estadísticos.

a) Normalidad

Vamos a comenzar el análisis comprobando y verificando si nuestro modelo cumple con la hipótesis de normalidad a través del análisis del gráfico QQ-plot donde veremos si los percentiles que presenta una muestra normal coinciden con los calculados en nuestro modelo. Para que cumpla el criterio de normalidad, los puntos analizados deben de estar situados sobre la diagonal; si no encontráramos esta situación, podría implicar una desviación de la normalidad.

Comprobaremos también el histograma de los residuos para compararlo con la campana de Gauss y buscar simetría con nuestro modelo. Si existiera desviación, cuanto mayor sea, menos se cumple la hipótesis de normalidad.

Gráfico 3: *Gráfico QQ-plot del modelo completo.*

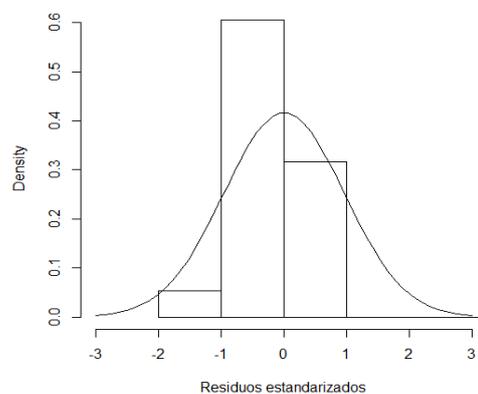


Fuente: Elaboración propia

Al analizar la figura arriba reproducida, correspondiente al gráfico QQ-plot, lo que buscamos es que la mayor parte de puntos que aparecen estén sobre la línea recta, lo cual sucede, lo único que en las dos colas hay valores que se alejan por lo que pueden surgir dudas de que se verifique la normalidad.

Comprobamos ahora, a través del histograma de residuos y su comparación con la campana de Gauss, si se cumple la hipótesis de la normalidad.

Gráfico 4: *Histograma de residuos del modelo completo.*



Fuente: Elaboración propia

Se observa que el histograma de nuestro modelo representado en el gráfico 4 no coincide con la forma de la campana de Gauss puesto que presenta una asimetría clara.

Respecto al cumplimiento de la normalidad, saldremos de dudas realizando el contraste de hipótesis de Shapiro-Wilks.

Tabla 6: Test de normalidad Shapiro-Wilks del modelo completo

Contraste de Shapiro-Wilks	
W	p-value
0.63138	0.00000001532

Fuente: Elaboración propia

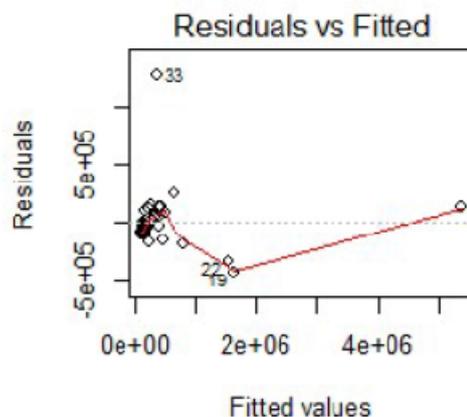
$H_0: e \sim \text{Normal}$
 $H_1: \text{Lo contrario}$ } p-value: = 0.00000001532 < 0.05

Este valor nos indica que hay evidencias en los datos para rechazar H_0 a favor de H_1 , por lo que rechazaremos la hipótesis de normalidad.

b) Homocedasticidad

Vamos a comprobar si nuestro modelo cumple la hipótesis de homocedasticidad o de varianza constante y por lo tanto está bien especificado. Para ello comenzamos analizando el gráfico de los residuos versus valores ajustados de la variable respuesta.

Gráfico 5: Gráfico de residuos vs. valores ajustados del modelo completo



Fuente: Elaboración propia

En el gráfico 5 queremos comprobar si la anchura de los puntos a lo largo de todo el gráfico es constante. Gráficamente deja dudas y no podemos afirmar que se cumpla la hipótesis de homocedasticidad. Para poder tener una idea más clara realizamos el test de Breusch-Pagan:

Tabla 7: Test de homocedasticidad de Breusch-Pagan del modelo completo

Breusch-Pagan test		
BP	df	p-value
5.0258	3	0.1699

Fuente: Elaboración propia

$$\left. \begin{array}{l} H_0: \text{Homocedasticidad} \\ H_1: \text{Heterocedasticidad} \end{array} \right\} \quad p\text{-value:} = 0.1699 > 0.05$$

Según los datos de la tabla 7 podemos decir que hay evidencias estadísticas en los datos para no rechazar H_0 a favor de H_1 . Por lo que parece que se cumple la hipótesis de homocedasticidad a pesar de las dudas surgidas en el análisis gráfico. Con esto podemos afirmar que nuestro

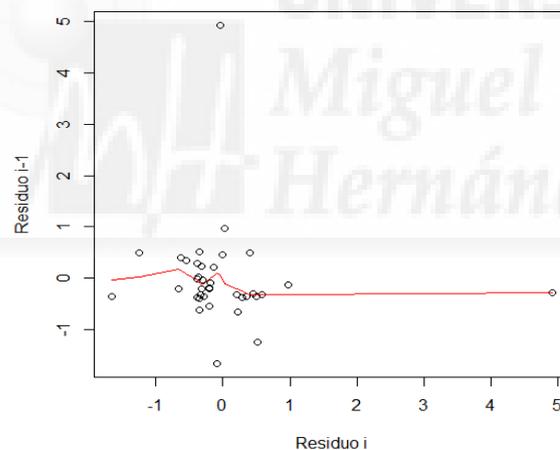
modelo, según este criterio, está bien especificado.

c) Incorrelación

Con la incorrelación lo que queremos demostrar es si los valores de unas variables pueden afectar a valores de otras variables. Una vez más realizaremos un análisis gráfico a través del gráfico de residuos i frente a residuos $i-1$ y una parte numérica en la que aplicaremos el test de Durbin-Watson.

Gráficamente:

Gráfico 6: Gráfico de incorrelación del modelo completo



Fuente: Elaboración propia

Gráficamente se observa incorrelación entre los residuos. No se ve ningunatendencia clara y continua creciente o decreciente en nuestros datos, por lo que podemos decir que se cumple la tercera hipótesis, aunque para confirmarlo vamos a realizar el test de Durbin-Watson.

Tabla 8: Test de incorrelación de Durbin-Watson del modelo completo

Durbin-Watson test		
DW	p-value	Alternative hypothesis
2.1409	0.748	true autocorrelation is not 0

Fuente: Elaboración propia

$$\left. \begin{array}{l} H_0: \rho = 0 \\ H_1: \rho \neq 0 \end{array} \right\} \quad p\text{-value:} = 0.748 > 0.05$$

El resultado que nos proporciona el test de Durbin-Watson, como podemos observar en la tabla 7, nos indica un p-valor=0.748 > 0.05, con lo que no se puede rechazar la hipótesis de incorrelación, no se puede rechazar H_0 a favor de H_1 .

d) La media de los errores es cero

Es la cuarta de las hipótesis básicas que debe cumplir nuestro modelo. Esta propiedad siempre se cumple como consecuencia de las herramientas que utilizamos ajustando el modelo, por lo que la suma de los errores siempre será cero.

6.- Transformación del modelo

Como consecuencia de los problemas de normalidad que presenta nuestro modelo, vamos a realizar una transformación de la variable respuesta para estudiar si mejoran basándonos en un método sencillo pero que a la vez sea aceptable.

Para ello, nos apoyaremos en un método que es el más utilizado para resolver los problemas planteados de normalidad y que pertenece a la familia de transformaciones potenciales Box-Cox.

Consiste en la introducción de un nuevo parámetro (λ) que nos ayude a

encontrar la estimación óptima y corrija los problemas de no linealidad. El valor del λ óptimo es aquel que, en nuestro modelo, explica el máximo de variabilidad. Esta transformación sólo es aplicable cuando la respuesta es positiva. Si hay valores cero o negativos, debemos proceder añadiendo una constante de forma que todos los valores sean positivos.

Para la serie original X de longitud N

$$X = x_0, x_1, x_2, \dots, x_{N-1}$$

La transformación Box-Cox de un parámetro es la siguiente:

$$x_i(\lambda) = \begin{cases} \frac{x_i^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \ln(x_i), & \lambda = 0 \end{cases}$$

donde $i = 0, 1, 2, \dots, N - 1$.

En función del valor de lambda, la transformación de Box-Cox incluye los siguientes casos especiales:

$$\begin{aligned} \lambda = -1.0, & \quad x_i(\lambda) = \frac{1}{x_i} \\ \lambda = -0.5, & \quad x_i(\lambda) = \frac{1}{\sqrt{x_i}} \\ \lambda = 0.0, & \quad x_i(\lambda) = \ln(x_i) \\ \lambda = 0.5, & \quad x_i(\lambda) = \sqrt{x_i} \\ \lambda = 2.0, & \quad x_i(\lambda) = x_i^2 \end{aligned}$$

Aplicándolo a nuestro modelo obtenemos que $\lambda = 0.4$, por lo que, podemos realizar la transformación Box-Cox correspondiente. Cuanto más cercano sea el valor de λ a 1 mejor será nuestro modelo.

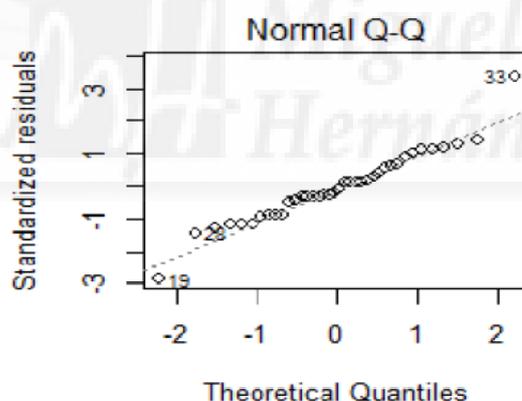
A partir de aquí, vamos a proceder al diagnóstico del modelo a través de un estudio gráfico y otro numérico en cada una de las hipótesis básicas de regresión lineal, que son:

- a) Normalidad
- b) Homocedasticidad
- c) Incorrelación
- d) La media de los errores es cero

a) Normalidad

Al igual que hemos hecho en el modelo original, comprobamos si nuestro modelo transformado cumple con la hipótesis de normalidad a través del análisis del gráfico QQ-plot y del histograma de los residuos.

Gráfico 7: Gráfico Q-Q Plot del modelo transformado



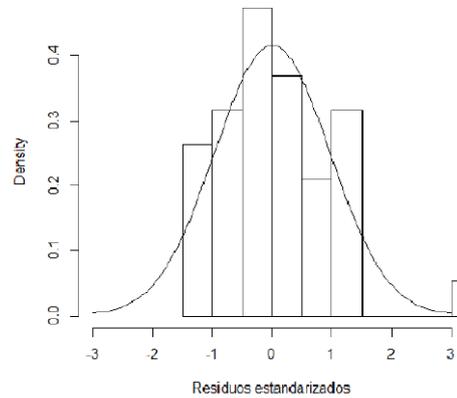
Fuente: Elaboración propia

Con el estudio del gráfico QQ-plot, observamos que sigue habiendo problemas de normalidad, ya que persiste la existencia de las dos colas, a pesar de que los valores centrales están mucho más concentrados. Pueden persistir los problemas de normalidad.

Comprobamos ahora el histograma de residuos para verificar si se cumple

la hipótesis de la normalidad.

Gráfico 8: *Histograma de residuos del modelo transformado.*



Fuente: Elaboración propia

Analizando el gráfico observamos que, si bien ha mejorado la distribución, sigue existiendo un problema de normalidad y presenta asimetría.

Con estos análisis, seguimos teniendo dudas acerca de que se cumpla la normalidad. Realizamos el contraste de hipótesis de Shapiro-Wilks.

Tabla 9: Test de normalidad Shapiro-Wilks del modelo transformado

Contraste de Shapiro-Wilks	
W	p-value
0.93816	0.03653

Fuente: Elaboración propia

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

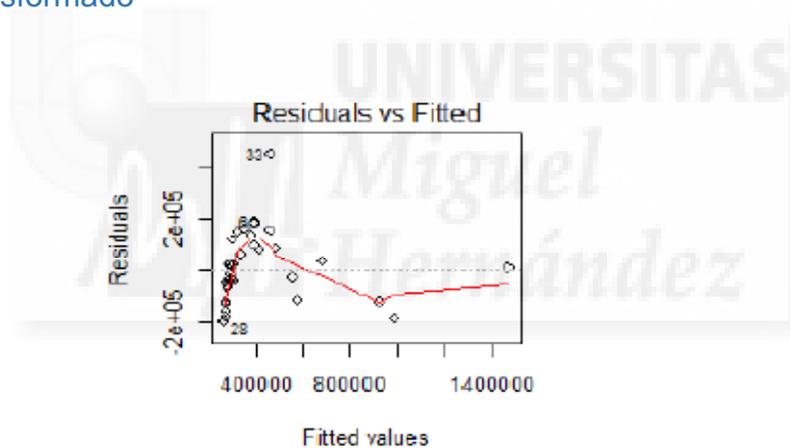
} p-value: = 0.03653 < 0.05

Este valor nos indica que hay evidencias en los datos para rechazar H_0 a favor de H_1 , por lo que rechazaremos la hipótesis de normalidad. Esto nos indica que no se cumple la normalidad a un nivel de error del 5%, aunque sí con niveles de error del orden del 3,5%.

b) Homocedasticidad

Vamos a comprobar si nuestro modelo transformado cumple la hipótesis de homocedasticidad o de varianza constante. Comenzamos analizando el gráfico de los residuos versus valores ajustados de la variable respuesta.

Gráfico 9: Gráfico de residuos estandarizados vs. valores ajustados del modelo transformado



Fuente: Elaboración propia

En el gráfico 9 no podemos observar que la anchura de los puntos a lo largo de todo el gráfico sea constante, por lo que no podemos afirmar que se cumpla la hipótesis de homocedasticidad. Para poder verificarlo con más seguridad, realizamos el test de Breusch-Pagan:

Tabla 10: Test de homocedasticidad de Breusch-Pagan del modelo transformado

Breusch-Pagan test		
BP	df	p-value
5.5439	3	0.136

Fuente: Elaboración propia

H_0 : Homocedasticidad }
 H_1 : Heterocedasticidad } p-value: = 0.136 > 0.05

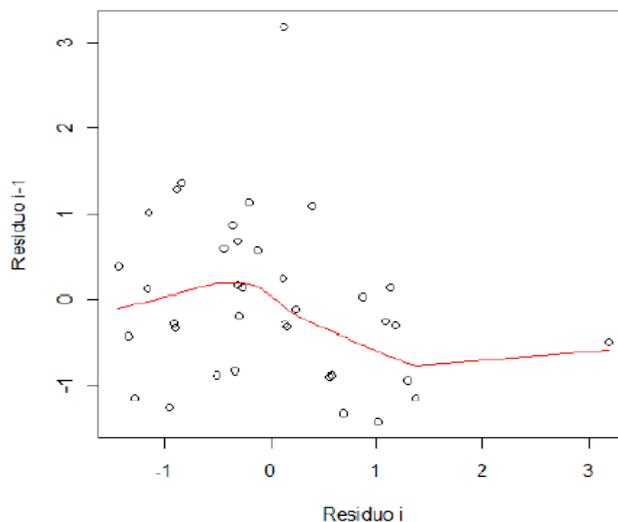
Con los datos obtenidos en la tabla 10 observamos que hay evidencias estadísticas en los datos para no rechazar H_0 a favor de H_1 por lo que se cumple la hipótesis de homocedasticidad a pesar de las dudas surgidas en el análisis gráfico. Con esto podemos afirmar que nuestro modelo transformado, según este criterio, está bien especificado.

c) Incorrelación

Vamos a estudiar si los valores de unas variables pueden afectar a valores de otras variables. Para ello analizamos la parte gráfica a través del gráfico de residuos i frente a residuos $i-1$ y la parte numérica a través del test de Durbin-Watson.

Gráficamente:

Gráfico 10: Gráfico de incorrelación del modelo transformado



Fuente: Elaboración propia

En el gráfico 10 se observa incorrelación entre los residuos porque no hay ninguna tendencia clara y continua, creciente o decreciente, en los datos por lo que podemos decir que se cumple la tercera hipótesis, aunque, para confirmarlo, vamos a realizar el test de Durbin-Watson. Para ello realizamos el contraste de hipótesis correspondiente.

Tabla 11: Test de incorrelación de Durbin-Watson modelo transformado

Durbin-Watson test		
DW	p-value	Alternative hypothesis
2.4415	0.2055	true autocorrelation is not 0

Fuente: Elaboración propia

$$\left. \begin{array}{l} H_0: \rho = 0 \\ H_1: \rho \neq 0 \end{array} \right\} \quad p\text{-value:} = 0.2055 > 0.05$$

El resultado del test de Durbin-Watson ($p\text{-valor}=0.2055 > 0.05$) al igual que observábamos en la tabla X, nos indica que no se puede rechazar la

hipótesis de incorrelación, no se puede rechazar H_0 a favor de H_1 , por lo tanto, se verifica esta hipótesis y se cumple la propiedad de incorrelación.

d) La media de los errores es cero

Es la cuarta hipótesis básica que debe cumplir nuestro modelo. Esta propiedad siempre se cumple como consecuencia de las herramientas que utilizamos ajustando el modelo por lo que la suma de los errores siempre será cero, cumpliendo el criterio de máxima verosimilitud.

7.- Procedimiento secuencial de selección de variables

A través del criterio AIC queremos estudiar si podemos eliminar, de nuestro modelo, alguna de las variables explicativas si resultan irrelevantes, con el fin de buscar el modelo más sencillo posible. Queremos encontrar el valor AIC más pequeño, pues es preferido a valores más altos. Este test nos proporcionará valor para el modelo completo y valor para los nuevos modelos en el caso de eliminar alguna de las variables explicativas. Los resultados obtenidos son:

Tabla 12: Tabla del criterio AIC

Variables		Valor AIC
Modelo completo inicial		951.35
- Number_of_employees		953.14
- Capital_stock		962.06
- Gross_value		980.03
Parámetros		
β_0	98569.1780	
β_1	-183.4	
β_2	2678	
β_3	957.4	

Fuente: Elaboración propia

Con los datos obtenidos, una vez aplicado el criterio AIC, podemos decir que el modelo que presenta un menor valor es nuestro modelo al completo con lo que no podemos eliminar ninguna covariable porque todas resultan necesarias para el modelo. Este hecho contrasta con los p-valores determinados en nuestro análisis previo sobre los coeficientes en el modelo asociado a cada variable explicativa. En concreto, la variable asociada al número de trabajadores resultaba no significativa en presencia de las otras dos covariables. Sin embargo, un criterio basado en la función de verosimilitud, como es el AIC, indica todo lo contrario. Esto hace que no sea del todo evidente si merece la pena eliminar el número de trabajadores del modelo o no.

8.- Estimación de la respuesta media de cada país

En esta sección, haciendo uso de nuestro modelo de regresión, estimaremos el nivel de contaminación esperado en función de las cantidades de inputs de todos y cada uno de los países de la muestra utilizada. La estimación se realizará de forma puntual y por intervalo de confianza al nivel habitual del 95%.

Hemos obtenido:

Tabla 13: Intervalos de confianza de la estimación de la respuesta media

Country	Emissions_of_CO2	2,50%	97,50%
Australia	402300,243	160.710,079	342.454,000
Austria	74683,1	62.869,157	253.499,800
Belgium	131355,91	87.067,399	279.057,800
Bulgaria	52077,22	2.253,418	208.752,100
Brazil	310158,7792	262.941,466	640.969,600
Canada	552329,0359	373.711,587	551.604,600
Cyprus	7978,54	-2.161,873	207.723,600
Czech_Republic	118297,47	18.268,299	208.070,800
Germany	898557,2	372.606,546	920.764,100
Denmark	85473,39	22.587,398	216.156,600
Spain	364888,73	205.952,317	414.450,500
Estonia	16686,07	-2.984,693	206.122,500
Finland	62272,92	44.360,045	246.131,100
France	427363,18	151.152,014	588.255,600
United_Kingdom	621585,19	695.596,166	894.819,500
Greece	113168,23	65.465,712	265.677,800
Hungary	61846,92	12.066,071	208.809,000
Indonesia	350116,1861	259.209,989	486.291,200
India	1197919,873	1.160.451,564	2.096.188,400
Ireland	47512,12	35.773,598	241.192,200
Italy	499357,6	217.258,940	512.959,000
Japan	1206901,409	1.184.416,536	1.879.137,800
Korea_Republic_of	533906,3358	314.732,026	491.886,400
Lithuania	16095,91	5.978,524	212.674,400
Luembourg	5356,5	2.121,140	211.880,200
Latvia	8564,58	1.900,892	211.015,400
Mexico	413758,5446	243.530,009	437.140,600
Malta	2608,25	-4.529,606	206.264,500
Netherlands	206417,06	118.285,906	301.814,500
Poland	324045,2	113.706,361	299.232,000
Portugal	71780,4	51.881,453	247.367,900
Romania	106363,27	16.216,510	217.358,700
Russia	1635392,026	185.282,114	526.341,800
Slovak_Republic	41370,45	12.338,079	216.435,300
Slovenia	17406,35	4.674,309	214.177,100
Sweden	62685,76	120.713,210	330.270,000
Turkey	248898,0898	44.476,395	242.743,500
United_States	5529034,166	4.881.052,634	5.915.402,900

Fuente: Elaboración propia

En la tabla 13, de color rojo, destacan aquellos países cuyo dato de emisiones de CO₂ se encuentra fuera del intervalo de confianza obtenido a través de la estimación de la respuesta media por arriba. Es decir, el resultado que arrojan estos países queda por encima del límite superior del intervalo de confianza obtenido al 95%. Estos países son: Australia, Canadá, República de Corea, Polonia, Rusia y Turquía.

Por el contrario, aquellos enumerados en color verde, son los países cuyo dato de emisiones de CO₂ no alcanza el límite inferior del intervalo de confianza obtenido a través de la estimación de la respuesta media. Estos países son: Reino Unido y Suecia.

El resto de los países, arrojan un dato que se encuentra dentro del intervalo de confianza.

Como conclusión, cabe destacar, que existen ciertos países que contaminan más de lo esperado, dados sus niveles de inputs, y, sin embargo, otros, menos de lo esperado.

9.- Conclusiones

En este último apartado del estudio econométrico vamos a exponer las principales conclusiones obtenidas de nuestra base de datos en relación con nuestra variable respuesta, la contaminación en el aire, y sus diversas magnitudes macroeconómicas básicas como el stock de capital, el número de trabajadores (en miles) y el valor añadido.

Para comenzar, se lleva a cabo un análisis preliminar a través de los dos métodos, gráfico (gráfico de dispersión) y numérico (correlaciones parciales y simples), donde se comprueba que existe una correlación muy alta entre nuestra variable respuesta, las emisiones de CO₂, y el valor añadido tomando como base precios del año 1995 y entre nuestra variable respuesta y el stock de capital. Con respecto al número de trabajadores (en miles) también existe esta correlación, aunque es menos elevada.

En segundo lugar, tras el análisis preliminar se procede a realizar el ajuste

del modelo para estimar el valor del intercepto y los coeficientes asociados a las tres covariables del modelo completo, y sus intervalos de confianza al 95%. Con estos resultados obtenidos, estudiamos sus p-valores para resolver el contrato de hipótesis respecto a si el coeficiente de dicha covariable puede ser nula, lo que indicaría que desaparecería de nuestro modelo.

En este primer análisis de los p-valores podemos concluir que las variables Capital_stock y Gross_value son necesarias para el modelo y no podemos eliminarlas, en presencia del resto de covariables y que no es significativa para nuestro modelo la variable Number_of_employees, lo que significa que, cuando están las otras variables en el modelo, ésta no añade información nueva y significativa sobre la variable respuesta (contaminación).

En tercer lugar, hemos procedido a realizar la bondad del ajuste para comprobar si se verifica que el modelo es bueno para explicar la variable respuesta a través de sus tres criterios: el estudio del error estándar residual, la tabla de ANOVA y el análisis del coeficiente de determinación.

En el estudio de error estándar residual hemos concluido que el valor es elevado, por lo que debemos decir que nuestro modelo no aprueba el primer criterio de la bondad del ajuste.

Así, con respecto al análisis de la varianza o tabla de ANOVA hemos cuantificado las diferencias entre la variabilidad de los datos y se obtiene un buen resultado. Nuestro modelo aprueba el criterio y es capaz de explicar la variabilidad contenida en nuestros datos.

Por último, hemos analizado el coeficiente de determinación o R^2 y R^2 ajustado que nos indica, en tanto por cien, la variabilidad total de los datos que explica el modelo de regresión. El resultado es un valor muy alto que nos permite afirmar que es un buen modelo y que ha sido capaz de explicar el 92% del valor de la variable respuesta.

Como resumen de la bondad del ajuste debemos decir que, nuestro modelo, no aprueba el criterio del error estándar residual, aunque sí que aprueba y con buena nota, los criterios de la tabla Anova y del coeficiente de

determinación.

El siguiente punto analizado ha sido el diagnóstico de las hipótesis básicas que nos presenta el modelo de regresión lineal en sus cuatro fases: normalidad, homocedasticidad, incorrelación y la media de los errores es cero.

Para estudiar la hipótesis de normalidad trabajamos con el análisis del gráfico QQ-plot y del histograma de residuos. Sus resultados son que tenemos dudas de que se cumpla la normalidad y se aplica el test de Shapiro-Wilks que su valor nos indica que hay evidencias en los datos para rechazar la hipótesis de normalidad.

Para la hipótesis de homocedasticidad o varianza constante se analiza, en primer lugar, el gráfico de los residuos versus valores ajustados de la variable respuesta donde no se observa que la anchura de los puntos en todo el gráfico sea constante por lo que no podemos afirmar que se cumpla esta hipótesis lo que nos lleva a aplicar, en segundo lugar, el test de Breusch-Pagan. Con los datos obtenidos podemos afirmar que se cumple la hipótesis de homocedasticidad.

Para la hipótesis de incorrelación se vuelve a realizar un análisis gráfico (el gráfico de residuos i frente a residuos $i-1$) donde no se aprecia ninguna tendencia clara y continua, por lo que se cumple la hipótesis de incorrelación, y, para confirmarlo, una parte numérica (el test de Durbin-Watson) donde no podemos rechazar H_0 a favor de H_1 , lo que indica que se vuelve a cumplir la hipótesis de incorrelación.

La cuarta hipótesis básica del modelo, la media de los errores es cero, se cumple siempre.

Como nuestro modelo no cumple la hipótesis de normalidad se aplica el método Box-Cox para intentar solucionar los problemas encontrados y se vuelven a aplicar las hipótesis básicas de normalidad, homocedasticidad, incorrelación y media de los errores cero.

Con la hipótesis de normalidad se vuelve a aplicar la parte gráfica, y aunque mejora la situación, se sigue teniendo dudas acerca de que se

cumpla la hipótesis de normalidad. Para la parte numérica se utiliza el test de Shapiro-Wilks que sigue evidenciando la falta de normalidad.

Para la hipótesis de homocedasticidad o varianza constante se analiza, en primer lugar, el gráfico de los residuos versus valores ajustados de la variable respuesta del modelo transformado donde no podemos afirmar que se cumpla esta hipótesis y para asegurarlo se aplica, en segundo lugar, el test de Breusch-Pagan donde si podemos afirmar que se cumple la hipótesis de homocedasticidad.

Para la hipótesis de incorrelación se analiza el gráfico de residuos i frente a residuos $i-1$ del modelo transformado donde se cumple la hipótesis de incorrelación, y, para confirmarlo, se realiza el test de Durbin-Watson que confirma que se cumple la hipótesis de incorrelación.

La cuarta hipótesis básica del modelo, la media de los errores es cero, ya hemos dicho que se verifica siempre.

Como conclusión de la transformación del modelo se ha de destacar que, tras la aplicación Box-Cox, mejora la hipótesis básica de normalidad que se presentaba, aunque no en la medida necesaria para poder ser no rechazada con un nivel de error del 5% aunque sí con niveles de error del orden del 3,5%. Esto significa que, realmente, no podemos confiar en la precisión de nuestros intervalos de confianza y p-valores a lo largo de todo el análisis desarrollado, pero sí en nuestras estimaciones puntuales, que se encuentran basadas en el principio de Mínimos Cuadrados.

El penúltimo paso llevado a cabo en el estudio ha sido la aplicación del criterio AIC para comprobar si alguna de las variables explicativas fuera irrelevante, pero, una vez aplicado el resultado ha sido que no se puede eliminar ninguna covariable porque todas son necesarias para el modelo.

Y, como último paso, se ha procedido a estudiar la respuesta media y obtener el intervalo de confianza al 95% para identificar en qué situación se encuentra cada uno de los países pertenecientes a nuestra base de datos con respecto a la contaminación, realizando la clasificación en países pertenecientes al intervalo de confianza, los que contaminan más de lo

esperado teniendo en cuenta sus niveles de inputs (Australia, Canadá, República de Corea, Polonia, Rusia y Turquía) y otros que contaminan menos de lo esperado (Reino Unido y Suecia).

10.- Bibliografía

1.-Revista Desarrollo Local Sostenible. Grupo Eumed.net y Red Académica Iberoamericana Local Global. Vol. 3, N.º 8. Recuperado el 27 de mayo de 2018 de www.eumed.net/rev/delos/08/

2.- Revista de Economía Industrial N.º 389. Recuperado el 3 de mayo de 2018 de <http://www.minetad.gob.es/Publicaciones/Publicacionesperiodicas/EconomiaIndustrial/RevistaEconomiaIndustrial/389/Jesús%20Iglesias.pdf>

3.- Aparicio, J., Martínez Mayoral, M., & Morales, J. (s.f.). Modelos Lineales Aplicados en R. (2004) (C. I. Hernández, Ed.) Dto. Estadística, Matemáticas e Informática.

4.- Aparicio, J. (s.f.). Econometría. Recuperado el 5 de febrero de 2017 de <http://umh3067.edu.umh.es/material/videos/>

5.- The R Project for Statistical Computing. Licencia GNU. Obtenido el 4 de marzo de 2017 de: <https://www.r-project.org/>

6.-VV.AA. Manual del Módulo de Sensibilización Ambiental. Edita Ministerio de Trabajo y Asuntos Sociales y Ministerio de Medio Ambiente. ISBN: 84-8320-217-4. Recuperado el 08 de mayo de 2018 de <http://www.mapama.gob.es/es/calidad-y-evaluacion-ambiental/temas/red-de-autoridades-ambientales-raa-/sensibilizacion-medioambiental/>

7.-Aparicio J. y otros. Journal of Environmental Management 196 (2017) páginas 148 – 160.

8.-<http://www.mapama.gob.es/es/>

9.-Economic growth and environmental quality: time series and cross-country evidence. Recuperado el 3 de mayo de 2018 de <http://documentos.bancomundial.org/curated/es/833431468739515725/pdf/multi-page.pdf>

10.- Cuadernos de economía (2012) 35, 43-54. Recuperado el 3 de mayo de 2018 de https://repositorio.uam.es/bitstream/handle/10486/681900/CE_97_%206.pdf?sequence=1&isAllowed=y

11.-ROCA, J. y PADILLA, E. (2003): «Emisiones atmosféricas y crecimiento económico en España: la curva de Kuznets ambiental y el protocolo de Kyoto». Economía industrial, nº 351, 2003. Recuperado el 3 de mayo de 2018 de <http://www.mincotur.gob.es/Publicaciones/Publicacionesperiodicas/EconomiaIndustrial/RevistaEconomiaIndustrial/351/Economia05.pdf>

12.- RUBIO, M. (2005).: «Economía, energía y CO₂: España 1850- 2000». Cuadernos Económicos de ICE, nº 70, pp. 51-71. Recuperado el 3 de mayo de 2018 de http://www.revistasice.com/CachePDF/CICE_70_51-76_827501E0F706BD313414025C5DBD011B.pdf

13.- ROCA, J. y ALCÁNTARA, V., (2001): «Energy intensity, CO₂ emissions and the environmental Kuznets curve. The Spanish case». Energy Policy, vol. 29(7), pp. 553-556. Recuperado el 3 de mayo de 2018 de <https://www.sciencedirect.com/science/article/pii/S0301421500001543>

14.- CARPINTERO, O. (2005): «El Metabolismo de la Economía Española: Recursos naturales y huella ecológica (1955-2000)».Madrid: Fundación César Manrique.Recuperado el 3 de mayo de 2018 de <http://www.fcmanrique.org/recursos/publicacion/elmetabolismo.pdf>

15.- RAMOS MARTÍN, J. (2003): «Intensidad energética de la economía española: una perspectiva integrada» UHE Workingpapers 2003_08, Universitat Autònoma de Barcelona, Departament d'Economia i Història Econòmica, Unitat d'Història Econòmica.

16.- DÍAZ-VÁZQUEZ, M. R. y CANCELO, M. (2010): «Análisis de los factores determinantes de la evolución de las emisiones de CO₂ y de azufre en países OCDE mediante una descomposición econométrica». Revista de Economía Mundial 26, 2010, pp. 85-106. Recuperado el 3 de mayo de 2018 de <http://www.redalyc.org/pdf/866/86618030004.pdf>

17.- JOSÉ MARCOS ALFAYATE Y OTROS (2008): «Contaminación ambiental. Una perspectiva desde la química». Editorial Rústica.

18.- JOHAN TELLI. «100 gestos para salvar el mundo». Art Books ediciones, SAS. (2013).

19.- JOANNA YARROW. «Eco ¡Lógico!: ¡Únete al debate medioambiental!» Editorial Blume. (2010).

