

ÍNDICE

| | |
|---|-----------|
| I. INTRODUCCIÓN | 4 |
| 1. La Emigración en España | 5 |
| 1.1. Concepto de Emigración | 5 |
| 1.2. Historia de la Emigración en España | 5 |
| 1.3. Noticias de Actualidad sobre la Emigración | 7 |
| 1.3.1. Portales de opinión en internet y movimientos en las redes sociales..... | 8 |
| 1.4. Estudios Recientes sobre la Emigración Española..... | 11 |
| 2. Formación requerida | 12 |
| II. OBJETIVOS..... | 13 |
| III. METODOLOGÍA Y TÉCNICAS | 17 |
| 1. Búsqueda | 18 |
| 2. Análisis Bivariante..... | 19 |
| 2.1. Concepto de Multicolinealidad..... | 19 |
| 2.1.1. Identificación..... | 20 |
| 2.1.2. Índice de acondicionamiento | 21 |
| 2.1.3. Solución | 21 |
| 3. Regresión Múltiple..... | 22 |
| 3.1. Concepto de Regresión Múltiple lineal y ecuación general..... | 22 |
| 3.2. Estimación de los parámetros y validación del modelo..... | 23 |
| 3.2.1. Significación..... | 23 |
| 3.2.2. Análisis Grafico..... | 24 |
| 3.2.3. Normalidad de los Residuos..... | 27 |
| 3.2.4. Heterocedasticidad | 27 |
| 3.2.5. Autocorrelación..... | 29 |
| 3.2.6. Puntos Atípicos e Influyentes | 31 |
| 3.3. Predicción | 32 |
| 4. Herramientas de cálculo | 32 |

| | |
|--|-----------|
| IV. CÁLCULOS Y RESULTADOS | 34 |
| 1. Matriz de Datos | 35 |
| 1.1. Análisis Bivariante (Multicolinealidad) | 36 |
| 1.1.1. Identificación..... | 36 |
| 1.1.2. Solución y comprobación | 39 |
| 2. Regresión Múltiple Lineal | 40 |
| 2.1. Modelo teórico | 40 |
| 2.2. Estimación de la primera ecuación de Regresión Múltiple y validación del modelo. | 41 |
| 2.2.1. Análisis de la Significación | 41 |
| 2.2.2. Análisis de los gráficos de residuos. | 42 |
| 2.2.3. Estudio de la Normalidad de los Residuos..... | 47 |
| 2.2.4. Heterocedasticidad | 48 |
| 2.2.5. Autocorrelación..... | 49 |
| 2.2.6. Residuos Atípicos y Puntos Influyentes. | 50 |
| 2.2.7. Solución problema Heterocedasticidad. | 51 |
| 2.2.8. Comprobación | 55 |
| 3. Modelo Definitivo..... | 58 |
| 3.1. Interpretación de los parámetros y predicción..... | 59 |
| 4. Planteamiento Segundo Modelo..... | 61 |
| 4.1. Modelo teórico. | 61 |
| 4.2. Estimación de la ecuación de regresión múltiple y validación del modelo. | 61 |
| 4.2.1. Análisis de la significación..... | 61 |
| 4.2.2. Análisis de los gráficos de residuos | 63 |
| 4.2.3. Estudio de la normalidad de los residuos..... | 65 |
| 4.2.4. Residuos atípicos y puntos influyentes | 65 |
| 4.3. Modelo teórico sin las comunidades de Cataluña, Comunidad Valenciana y Madrid. | 66 |
| 4.3.1. Análisis de la significación..... | 66 |
| 4.3.2. Estudio de la normalidad de los residuos..... | 67 |
| 4.3.3. Heterocedasticidad | 68 |
| 4.3.4. Autocorrelación..... | 68 |
| 4.3.5. Residuos atípicos y puntos influyentes | 69 |
| 4.3.6. Interpretación de los parámetros y Predicción | 69 |

| | |
|--|-----------|
| V. CONCLUSIONES | 71 |
| VI. FUTURAS LÍNEAS DE INVESTIGACIÓN | 75 |
| VII. BIBLIOGRAFÍA | 79 |

I. Introducción

1. La Emigración en España

1.1. Concepto de Emigración

La emigración es un movimiento de la población desplazándose desde el lugar de origen o residencia para establecerse en otro país o región, normalmente por causas económicas o sociales desfavorables. Este movimiento no es exclusivo de los países subdesarrollados, sino que está presente en mayor o menor medida en casi todos los países, aunque es una característica más propia de los países pobres, cuya población en edad de trabajar intenta mejorar sus condiciones de vida [1].

Tal y como se ha mencionado, las razones por las que generalmente la población deja sus países de origen suelen ser problemas económicos, aunque en realidad los individuos toman esta decisión por otras y muy complejas circunstancias. Entre las más recurrentes se pueden nombrar los problemas referentes al nivel de vida (Problemas de desempleo, bajos salarios y poco reconocimiento profesional) por razones políticas, persecuciones dentro del país (razones raciales, políticas y religiosas), Guerras civiles o razones medioambientales (catástrofes naturales como huracanes o terremotos) [2].

La palabra emigración va ligada de manera recíproca a la palabra inmigración. Para empezar hay que partir de la palabra migración que define los movimientos de personas de un lugar a otro. Como hemos tratado anteriormente en la emigración las personas salen de su país para ir a otro distinto. El País que recibe a estas personas denomina a estos como inmigrantes. Por tanto, toda aquella persona que abandona su país de origen es tanto emigrante del país del que viene, como inmigrante al país de destino [3].

Hay un frase de Isabel Allende que dice: “Aprendí pronto que al emigrar se pierden las muletas que han servido de sostén hasta entonces, hay que comenzar desde cero, porque el pasado se borra de un plumazo y a nadie le importa de dónde uno viene o que ha hecho antes.”

1.2. Historia de la Emigración en España

La historia de España está marcada por el fenómeno de la emigración. Esta emigración determinó la fragmentación del país y de su gente, dividiendo las familias, los afectos, los roles, las geografías, los hábitos y los entornos culturales [4]. Se pueden diferenciar tres periodos principales dentro de la historia de España como emisor de emigrantes:

- 1830-1945: Época caracterizada por las migraciones hacia América Latina.

Son varios los factores que influyeron en esta salida masiva de población española. El predominio de la pequeña propiedad, un sistema hereditario desigual, altas tasas de analfabetismo, bajos niveles salariales y un escaso desarrollo urbano e industrial; pero, fundamentalmente, destaca la crisis agropecuaria por la que estaba pasando España en esos momentos. Por esta razón, muchos campesinos se vieron obligados a abandonar las áreas rurales, pero se encontraron que las ciudades, en pleno proceso de industrialización, no podían absorber a tanta población, por lo que la única salida para muchos fue emigrar. Mención especial se merece el más de medio millón de personas que se vieron obligadas a abandonar España entre 1936 y 1939 debido a la Guerra Civil. Los países elegidos por los refugiados españoles fueron Francia, Rusia, México, Argentina y otros países latinoamericanos, fundamentalmente. La Guerra Civil supone la emigración forzosa de españoles más importante que se produce a lo largo de toda nuestra historia.

El gobierno del país fomentó la emigración hacia América a partir de 1853, y al final del siglo se habían ido medio millón de personas, cifra que en ese momento representaba el 20% de su población.

- 1945-1973: Migraciones hacia otros países europeos.

Los emigrantes españoles partían principalmente de Andalucía y Levante, debido a las altas tasas de desempleo. Los países elegidos como destino por nuestra población fueron, fundamentalmente, Alemania, Francia, Gran Bretaña, Holanda y Suiza. Los factores desencadenantes de este fenómeno guardan relación con el desarrollo económico de los países del norte, que demandaban mano de obra.

Como consecuencia de la Guerra Civil otro medio millón de personas salen del país. La corriente migratoria se reanuda a partir de 1940, y durante veinte años (hasta 1960) salen hacia América unos 32.000 ciudadanos al año, alcanzando una cifra superior a los 600.000. A partir de 1960 y hasta 1973, la emigración da un gran giro, dirigiéndose entonces los españoles hacia los países europeos.

- 1973-1985: Crisis del petróleo e inicio de restricciones en los movimientos migratorios intraeuropeos.

La crisis del petróleo de 1973 desencadena una inestabilidad económica que produce un cambio significativo en los flujos migratorios de todos los países. Así, comienza a definirse una tendencia caracterizada por las fuertes restricciones a la llegada de nuevos trabajadores, que se concretan fundamentalmente en dos medidas: Cierre de fronteras y fomento del retorno de trabajadores extranjeros a sus países de origen. Es un intento por frenar la llegada de mano de obra externa ante la aparición del problema del

desempleo en los países tradicionalmente receptores. En los años posteriores, la política de restricción permanece vigente, pero la inmigración continúa creciendo paulatinamente. Además, varios países del sur de Europa, entre ellos España, se convierten por vez primera en países receptores de inmigración. El número de extranjeros residentes legales en España empieza a aumentar en los años 80, momento en el que nuestro país consolida su democracia e inicia su proceso de integración en la Unión Europea [5].

1.3. Noticias de Actualidad sobre la Emigración

En este apartado se procede a explicar algunos artículos de opinión en los cuales se habla sobre la emigración de los españoles en los últimos tiempos.

El primero de los artículos que se quiere destacar fue escrito el 7 de noviembre de 2013, en el periódico el país. Este artículo se titula “El 89% de los españoles que emigran en busca de empleo cuenta con educación superior”, en él se explica que según un estudio realizado por el Real Instituto Elcano revela el 89% de los emigrantes españoles cuentan con educación superior. En el estudio se ha consultado a una muestra de más de 1.500 ciudadanos españoles que han salido de su país y se han tomado datos más exactos sobre la situación. En total, se ha obtenido respuesta de 7.070 ciudadanos de Grecia, Italia, Portugal y España. González ha manifestado que la mayoría de los emigrantes son hombres y que la profesión que mayormente busca oportunidades en otros sitios de Europa son los ingenieros. Las cifras dicen que un 70% de los que han optado por dejar España están trabajando, un 25 % no está ocupado, 11 % están buscando trabajo y un 6 % son estudiantes [6].

El segundo artículo a ser comentado fue escrito el 10 de octubre de 2013, por Pilar Almenar Nava, en el periódico “el país”. Este artículo se llama “Nos obligan a una vida muy precaria” y habla de cómo los jóvenes de entre 25-35 años se ven obligados vivir en unas pésimas condiciones laborales si se quedan en España. Se cuenta la historia de varios chicos/as que decidieron no abandonar este país, porque no querían o simplemente porque no podían económicamente. Porque para irse, también hace falta dinero. Algunos de los que se quedaron se quejan de que es imposible ser emprendedores y montar pequeñas empresas porque los impuestos son carísimos y no se tiene en cuenta los beneficios, otros hablan de trabajos precarios en los que su salario oscila entre los 540€, una chica en concreto con una licenciatura en arquitectura ha decidido montar una academia de repaso porque en su campo le es imposible encontrar trabajo. En resumen, los jóvenes que decidieron no emigrar fuera de España, ya tengan

formación o no, se sienten ninguneados y creen que sus trabajos no están lo suficientemente reconocidos [7].

1.3.1. Portales de opinión en internet y movimientos en las redes sociales.

Existen infinidad de portales en internet en los cuales se habla sobre la emigración, sobre como los emigrantes españoles dejan sus países de origen para buscar un futuro mejor. En este punto, se intenta hacer un resumen de los más interesantes.

MOVIMIENTOS Y CAMPAÑAS:

Un ejemplo claro de algunos de estos movimientos es una página web llamada: NO NOS VAMOS, NOS ECHAN (Figura I. 1). No nos vamos, nos echan, es una iniciativa creada por la Juventud sin futuro para visibilizar las verdaderas causas del exilio laboral de los jóvenes españoles y construir alternativas.



Figura I. 1. Cabecera de la página web: No nos vamos nos echan [8].

La reivindicación "No nos vamos, nos echan" es de Juventud Sin Futuro(JSF) (Figura I. 2), un colectivo de jóvenes de la órbita del 15-M, que en 2013 lanzó una campaña para denunciar la "situación de exilio forzoso de la juventud". Ante una tasa de paro juvenil de más del 50%. Los promotores de la iniciativa quieren denunciar que esta situación no es accidental, sino que "tiene culpables". Así, remarcan que España "no es país para jóvenes" ya que estos se ven obligados a elegir entre "la desocupación, la precariedad o el exilio forzado" [8]. En el mapa que se observa a continuación se muestra cuales son los destinos más poblados de emigrantes españoles.



Figura I. 2. Mapa de los lugares más poblados por emigrantes españoles [8].

PAGINAS WEB DE AYUDA PARA BUSCAR TRABAJO EN EL EXTRANJERO:

Existen algunas páginas web o blogs creados exclusivamente para orientar a la gente que busca trabajo en el extranjero, en estos portales se pueden encontrar orientación, ofertas de empleo e incluso formación [9], como puede verse en el portal del ejemplo de la Figura I. 3.



Figura I. 3. Página web para la orientación y búsqueda de empleo en el extranjero [9].

PROGRAMAS DE TELEVISIÓN:

Españoles en el mundo (Figura I. 4) es un programa de televisión española que refleja claramente como es la vida de algunos de los españoles que por diversos motivos han tenido que emigrar a otros países. En este programa se observan todo tipo de historias [10]. ¿Hasta qué punto los españoles emigran? Si se realiza un programa de televisión

para mostrar a la gente como viven los españoles emigrantes ¿será porque resulta ser un tema interesante?



Figura I. 4. Imagen del programa de televisión: españoles en el mundo [10].

REDES SOCIALES:

Existen todo tipo de grupos y comunidades en las redes sociales para aquellos emigrantes españoles que se desplazan a otro país. Estas páginas ofrecen ayuda y ponen en contacto a las personas para que se ayuden unos a otros e intercambien opiniones sobre su lugar de destino [11]. Como por ejemplo las que se pueden ver en la Figura I. 5.



Figura I. 5. Grupos de ayuda en las redes sociales (Facebook).

1.4. Estudios Recientes sobre la Emigración Española.

Hay muchos estudios referentes a la emigración española en los últimos tiempos. Cada uno distinto al anterior, entre ellos un estudio realizado por el instituto nacional de estadística en 2013 “Estadísticas de Migración 2013” [12]. En ese estudio se pueden observar datos, gráficos y tablas de cómo varía la población en España desde el 2008 hasta el 2013. Se intenta responder a cuáles son los motivos de estos descensos tan bruscos de la población, todo representado por Comunidades Autónomas. En ese estudio no se realiza una modelización o un intento de predicción del flujo de emigración, sino que se limita a un análisis descriptivo del incremento de la variable en el tiempo (Figura I. 6).

En este proyecto se pretende ir más allá complementado esta clase de estudios. Se va a proceder a realizar un análisis sobre los factores que afectan a la emigración Española en 2013, con el empleo de técnicas estadísticas de análisis multivariante, y el software *Statgraphics centurión VXI.II*. Con esto se intentará modelizar el comportamiento del flujo de emigración española y se analizará la posibilidad de poder realizar predicciones a corto plazo en función de las variables que se demuestre que explican el fenómeno de la emigración en España.

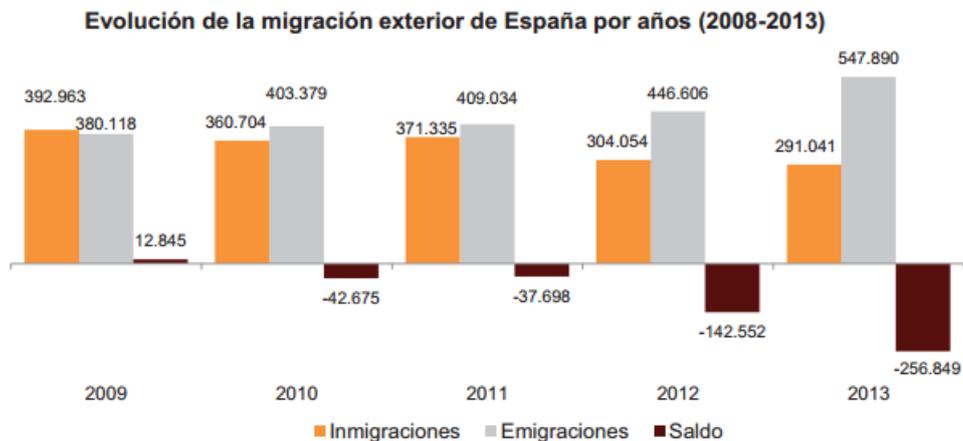


Figura I. 6. Gráfico de barras de la evolución de las migraciones de España 2008-2013 [12].

2. Formación requerida

Tanto para la búsqueda de información como para su tratamiento, el alumno ha necesitado los conocimientos adquiridos en las asignaturas:

- Econometría: Para el planteamiento, cálculo y validación de las series temporales obtenidas a partir de la matriz de datos generadas.
- Introducción a la estadística: Para la interpretación de la normalidad de los residuos mediante la campana de GAUSS y los distintos test de hipótesis.
- Métodos estadísticos en Economía: Para el planteamiento e interpretación de los test de hipótesis y la interpretación de la varianza de los residuos.
- Modelos Matemáticos para ADE: Para los cálculos realizados durante el proyecto.
- Investigación comercial: Para el planteamiento del proyecto y la investigación sobre el mismo.
- Economía Española: Para la obtención de los datos que se van a analizar en cada una de las regresiones múltiples.

II. Objetivos

En este apartado se plantean los objetivos del trabajo. El objetivo principal es determinar cuáles son los factores más relevantes en la emigración de los españoles por comunidades autónomas en los últimos tiempos, mediante la modelización del flujo de emigración, y poder realizar predicciones del mismo. Para conseguir el objetivo principal se plantean los siguientes objetivos secundarios:

- Obtener los datos necesarios para poder explicar cuáles son los factores más importantes en la emigración de los españoles hacia el extranjero.
- Realizar un análisis bivalente para determinar el grado de correlación entre las variables seleccionadas para el modelo y eliminar, si procede, aquellas que estén directamente correlacionadas.
- Plantear modelos de regresión múltiple con los que se pueda explicar el flujo de emigración de los españoles en función de las variables validadas en el análisis bivalente.
- Validar los modelos de regresión múltiple planteados y comprobar la robustez de los mismos mediante los contrastes de hipótesis, significatividad, normalidad, heterocedasticidad y autocorrelación.
- Realizar predicciones de los modelos validados y comprobar con los datos reales publicados posteriormente el acierto de estas predicciones.

III. Metodología y Técnicas

1. Búsqueda

Cuando se empezó este proyecto, se hizo una primera búsqueda en el INE (Instituto Nacional de Estadística) de variables que pudieran estar relacionadas o explicar los movimientos migratorios en España [13].

Las primeras variables que se encontraron fueron:

1. Becas de enseñanza obligatoria por comunidades autónomas (Becas_CCAA_OBLIG). Unidades: número de becas, ayudas y becarios. Importe en miles de euros.
2. Becas de enseñanza postobligatoria No Universitaria por comunidades autónomas (Becas_CCAA_PostNOUNI). Unidades: número de becas, ayudas y becarios. Importe en miles de euros.
3. Becas de enseñanza universitaria por comunidades autónomas (Becas_Univ). Unidades: número de becas, ayudas y becarios. Importe en miles de euros.
4. Becas del ministerio de educación en enseñanza obligatoria (Becas_MICIN_OBLIG). Unidades: número de becas, ayudas y becarios. Importe en miles de euros.
5. Becas del ministerio de educación en enseñanza postobligatoria No Univeritaria (Becas_MICIN_PostNOUNI). Unidades: número de becas, ayudas y becarios. Importe en miles de euros.
6. Becas del ministerio de educación en enseñanza Universitaria (Becas_MICIN_Univ). Unidades: número de becas, ayudas y becarios. Importe en miles de euros.
7. Producto interior bruto por CCAA (PIB). Unidades: Euros.
8. Índice de precios de Consumo por CCAA (IPC). Unidades: Índice, Tasas.
9. Índice de producción industrial por CCAA (IPI). Unidades: Índice, Tasas.
10. Tasa de Desempleo al finalizar la enseñanza obligatoria (Desempleo_ESO). Unidades: porcentajes: personas.
11. Inversión en I+D+I publica por CCAA (I+D+i_PUB). Unidades: Miles de euros.
12. Inversión en I+D+I privada por CCAA (I+D+i_Empresa). Unidades: Miles de euros.
13. Índice de Paro por CCAA (PARO). Unidades: Tasas.

14. Índice de Paro juvenil por CCAA (PARO_JUVENIL). Unidades: Miles de personas.
15. Índice de Precios de la Vivienda por CCAA (IPVIVIENDA). Unidades: Índice, Tasas.

2. Análisis Bivariante

Una de las hipótesis del modelo de regresión, referida a las variables explicativas, es que éstas no deben estar relacionadas entre sí, ya que esto puede afectar a la estimación de los parámetros del modelo, de forma que dependa del orden en el que se incluyan. Cuando esto sucede, se dice que el modelo presenta problemas de multicolinealidad.

2.1. Concepto de Multicolinealidad

Cuando dos o más variables explicativas están relacionadas entre sí, o toman valores semejantes (proporcionales) en la muestra observada, entonces se tiene una situación de multicolinealidad en el modelo.

En primer lugar la multicolinealidad se dará cuando existan relaciones lineales, exactas o no, entre las variables explicativas del modelo. En segundo lugar, la multicolinealidad se dará también en el caso de que exista una proporcionalidad entre los valores de las variables explicativas, sin haber realmente una relación entre ellas.

La multicolinealidad es un problema porque ésta afecta tanto a las estimaciones de los parámetros como a la varianza de la perturbación, y por lo tanto falsea las pruebas de hipótesis para comprobar si son significativos. Otro problema que se daría es que según el orden en el que se introduzcan las variables en el modelo las variables resultantes sean significativos o no de forma errónea.

2.1.1. Identificación

Existen tres pruebas para encontrar los distintos tipos de multicolinealidad posibles:

- Análisis de la matriz correlación Rii
- Análisis de la matriz inversa inversa Rii-1
- Cálculo del Índice de acondicionamiento

Es necesaria la realización de las tres pruebas, de forma que si una de las tres señala problema se debe aplicar una solución.

2.1.1.1. Matriz de Correlación

Con el fin de detectar un problema de relación entre las variables por parejas, la mejor referencia es el coeficiente de correlación Rii (1). La matriz de correlación R de las variables explicativas es una matriz simétrica de orden k cuyo término general ij es el coeficiente de correlación lineal simple entre las variables Xi y Xj.

$$R_{ij} = \frac{cov(X_i, X_j)}{\sqrt{var(X_i) var(X_j)}} \quad (1)$$

Por lo tanto la matriz R tiene unos en la diagonal principal, y el resto de los elementos son los coeficientes de correlación de cada pareja de variables explicativas. Los Rii de la Figura III. 1 corresponden a los coeficientes de correlación entre las variables.

| | x1 | x2 | x3 |
|----|-----------------|-----------------|-----------------|
| x1 | R ₁₁ | R ₁₂ | R ₁₃ |
| x2 | R ₂₁ | R ₂₂ | R ₂₃ |
| x3 | R ₃₁ | R ₃₂ | R ₃₃ |

Figura III. 1 Ejemplo de matriz de correlación.

El criterio que se emplea en este proyecto es el siguiente: En el caso de que $|R_{ii}| > 0.7$, significa que existe problema de Multicolinealidad entre dos variables.

2.1.1.2. Matriz de correlación Inversa

La matriz de correlación inversa (Figura III. 2), relaciona una variable explicativa con todas las demás. Para obtener esta matriz se realiza el cálculo con Excel. El procedimiento consiste en observar los elementos de la diagonal principal y si alguno supera el valor de 10 ($R_{ii-1} > 10$), entonces se confirmará que existe multicolinealidad entre una variable y el resto.

$$\begin{matrix} x1 \\ x3 \\ x2 \end{matrix} \begin{pmatrix} 21,76 & -21,25 & 0,38 \\ -21,25 & 21,75 & -0,36 \\ 0,38 & -0,36 & 1,01 \end{pmatrix}$$

Figura III. 2. Ejemplo matriz de correlación Inversa

2.1.2. Índice de acondicionamiento

Este método se basa en los autovalores de la matriz de correlación. Se define un índice de acondicionamiento (I.C) calculado según la ecuación (2). Si el I.C es mayor que 1 y menor que 10/20 significa que no existe multicolinealidad, en caso contrario habrá problemas de multicolinealidad moderada o elevada.

$$I.C. = \sqrt{\frac{\text{Autovalor}_{\max}}{\text{Autovalor}_{\min}}}$$

(2)

2.1.3. Solución

Para solucionar la multicolinealidad, independientemente del la clase que se haya detectado, se procede eliminando las variables que generan el problema. Se debe empezar siempre por la que obtenga siempre el valor más elevado en la matriz de

correlación inversa. Si en dos variables con problema de multicolinealidad los valores de la diagonal no difieren en más de una unidad, no sería correcto eliminar directamente una de esas dos variables. En este caso se deben plantear dos modelos de regresión múltiple, eliminando en cada uno una de las dos variables candidatas a eliminar. Se seleccionará el modelo que presente mayor ajuste (R^2).

3. Regresión Múltiple

A continuación se van explicar los fundamentos del modelo multivariante planteado en este proyecto.

3.1. Concepto de Regresión Múltiple lineal y ecuación general.

Mediante un modelo de regresión lineal múltiple se pretende explicar el comportamiento de una determinada variable que denominaremos variable a explicar, variable endógena o variable dependiente, (y representaremos con la letra Y) en función de un conjunto de variables explicativas X_1, X_2, \dots, X_i mediante una relación de dependencia lineal (3).

Y = Variable dependiente

$X_1, X_2 \dots X_i$ = Variables explicativas

$$Y = \beta_0 + \beta_1 \cdot X_1 + \dots + \beta_i \cdot X_i + U \quad (3)$$

β_0 = Valor medio de la variable Y cuando todas las variables explicativas valen 0.

β_i = Incremento medio de la variable Y cuando aumenta en 1 unidad X_i . Y el resto de las variables explicativas permanecen constantes.

U = La parte de los datos que no son capaces de explicar en el modelo, es decir, aquellas variables que no se han tenido en cuenta y que tienen efecto en la variable dependiente (También es el error).

3.2. Estimación de los parámetros y validación del modelo

3.2.1. Significación

En este apartado se va a explicar el contraste de hipótesis que permite determinar si son significativas las variables explicativas incluidas en el modelo de forma individual y si lo es modelo en sí.

El contraste del modelo determinara si todos los parámetros son igual a cero o no lo son. Así que la hipótesis sobre el modelo permite determinar si al menos uno de los parámetros del modelo es significativo (4) (excepto el termino independiente (β_0)). En cualquier contraste de hipótesis se acepta la hipótesis nula H_0 , siempre que el P-Valor sea mayor o igual que α (0.05).

$$\left. \begin{array}{l} H_0: \beta_1 = \beta_2 = \dots = 0 \\ H_1: \text{Algún } \beta_1 \neq 0 \end{array} \right\} \alpha = 0,05 \quad (4)$$

Si Aceptamos $H_0 \rightarrow$ Modelo NO significativo.

Si rechazamos H_0 y aceptamos $H_1 \rightarrow$ Modelo Significativo.

El contraste de hipótesis sobre un parámetro individual (5) permite determinar qué variables influyen realmente en la variable dependiente, y cuáles deben eliminarse, incluyendo la constante β_0 .

$$\left. \begin{array}{l} H_0: \beta_0 = 0 \\ H_1: \beta_0 \neq 0 \end{array} \right\} \quad \left. \begin{array}{l} H_0: \beta_1 = 0 \\ H_1: \beta_1 \neq 0 \end{array} \right\} \quad (5)$$

$$\alpha = 0,05$$

Si Aceptamos $H_0 \rightarrow$ Variable NO significativa.

Si rechazamos H_0 y aceptamos $H_1 \rightarrow$ Variable Significativa

La solución al problema de la significación consiste en la búsqueda de otras variables o el planteamiento de otras estructuras no lineales. En caso de que la constante no sea significativa, se procede a eliminar la contante. En cambio, si un parámetro no es significativo, se elimina la variable que acompaña al parámetro.

3.2.2. Análisis Grafico

Mediante el análisis gráfico de los residuos frente a la estimación y variables explicativas, se podrá apreciar la posibilidad de problemas de Heterocedasticidad, Autocorrelación de primer orden, Falta de linealidad y puntos anómalos.

En el gráfico de residuos frente a la estimación, se puede observar los siguientes problemas:

- Heterocedasticidad: esto ocurre cuando la varianza de U no es constante, de forma que aumenta con la predicción (Figura III. 3).

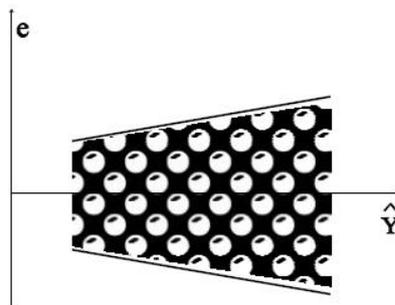


Figura III. 3. Ejemplo de un gráfico de residuo para un modelo con problemas de heterocedasticidad.

- Punto Anómalo: se observan cuando alguna de las perturbaciones se aleja mucho de la pauta marcada por el resto (Figura III. 4).

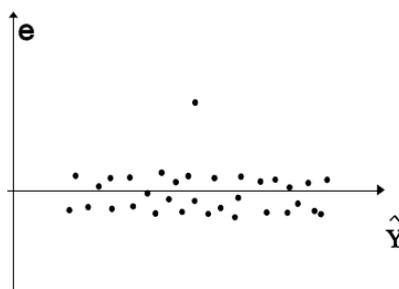


Figura III. 4. Ejemplo de residuos anómalos.

- Falta de linealidad: se presenta el caso de suponer una relación lineal cuando en realidad es cuadrática. Se suele detectar cuando los residuos se distribuyen en forma parabólica (Figura III. 5).



Figura III. 5. Ejemplo falta de linealidad.

La falta de linealidad es el único de los problemas que se detecta únicamente de forma gráfica y no en un contraste de hipótesis, por tanto se procederá a describir en este punto como se solucionaría este problema. Para solucionar este problema se debe elevar la variable explicativa que genera el problema al cuadrado tantas veces como sea necesario (6) hasta que deje de detectarse el problema. Se detienen las modificaciones cuando no se detecta la forma de parábola y el ajuste del modelo deja de aumentar.

$$X \rightarrow X^2 \rightarrow X^4 \rightarrow X^8 \rightarrow X^{16} \rightarrow X^{32} \quad (6)$$

$$X^{n^2} = X^{n*2}$$

- Falta del término constante: en el caso de que se haya planteado un modelo sin término independiente (β_0), y este deba estar presente, se observaría una ordenación ascendente de los residuos (Figura III. 6).



Figura III. 6. Gráfico del residuo de un ajuste sin contraste.

- Combinación de problemas: puede darse el caso de que se presente una combinación de los problemas descritos de forma que resulte difícil detectarlos (Figura III. 7).



Figura III. 7. Ejemplo de residuo con problemas de heterocedasticidad y falta de linealidad.

Por último, en el gráfico de los residuos frente al número de fila se pueden observar los siguientes problemas:

- Problemas de Autocorrelación de primer orden negativa: se detectan cuando los residuos se distribuyen de forma sinusoidal alternando valores positivos y negativos (Figura III. 8).

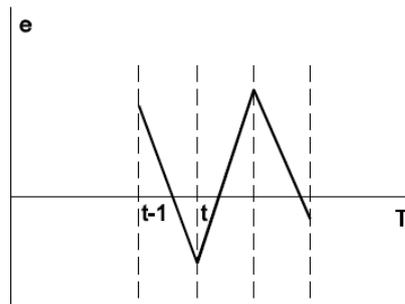


Figura III. 8. Ejemplo Autocorrelación primer orden Negativa.

- Autocorrelación de primer orden positiva: se detecta gráficamente cuando los residuos son positivos y con el tiempo hay una caída abrupta a valores negativos (Figura III. 9).

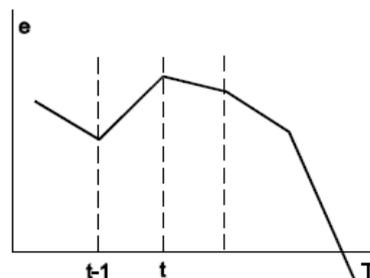


Figura III. 9. Ejemplo Autocorrelación primer orden Positiva.

3.2.3. Normalidad de los Residuos

Este contraste de hipótesis se realiza para comprobar que los residuos siguen una distribución normal. Si esto no ocurre se pueden dar problemas de ajuste en el modelo, y afectar a la estimación de los parámetros. El contraste de hipótesis es el siguiente:

- H_0 : Los residuos se distribuyen normalmente
- H_1 : Los residuos NO se distribuyen normalmente

El software plantea la posibilidad de realizar cuatro tipos de test para la normalidad que proporcionan los p-Valores para resolver el test de hipótesis. Se debe seleccionar el test más restrictivo (el que tiene el P-valor más pequeño), para confirmar que se acepta la normalidad de los residuos.

3.2.4. Heterocedasticidad

La heterocedasticidad es un problema que aparece cuando la varianza de la perturbación (U) no es constante. U debe tener una distribución normal de media cero y desviación Típica (σ^2) constante. Si σ^2 no es constante (Figura III. 10), tendremos un error en el cálculo de: estimación de parámetros, predicciones, y contrastes de hipótesis.

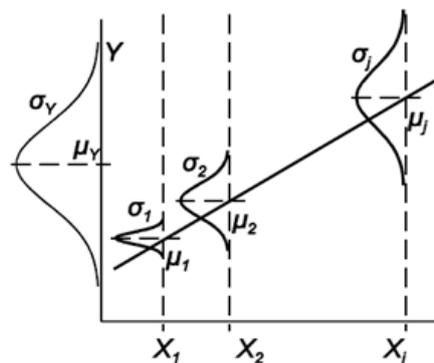


Figura III. 10. Representación de la varianza de la perturbación cuando se da el problema de Heterocedasticidad.

A continuación se enumeran algunas de las causas de la heterocedasticidad:

- Alguna de las variables tiene intrínseco el problema de heterocedasticidad.
- Que se introduzca en los datos debido a que los datos provengan del promediar un cierto número de datos.
- Se ha planteado mal el modelo.

- Que haya puntos anómalos entre los datos, esto puede provocar heterocedasticidad en el ajuste del modelo.

Este problema se puede identificar de dos formas:

- Mediante análisis gráfico→ Con gráficos que relacionen los residuos con la variable a explicar predicha, o frente a las variables explicativas en concreto para ver cuál es la causante. La identificación suele ser fácil debido a que la forma del gráfico es muy característica, en forma cónica (Figura III. 11).

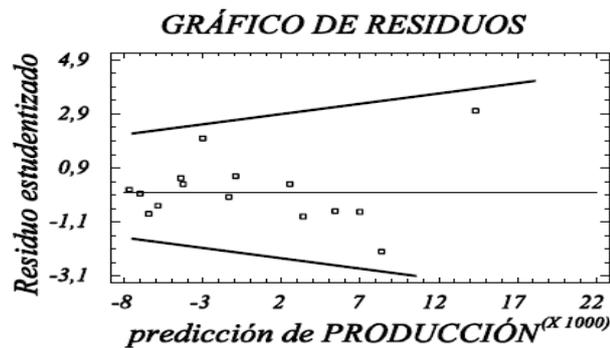


Figura III. 11. Grafico problema de Heterocedasticidad.

- Mediante contrastes de hipótesis→ Se debe que calcular un modelo lineal que relacione los residuos al cuadrado del modelo con las variables explicativas (7). De manera que si alguno de los parámetros, beta que no sea la constante (β_0), fuera significativo, se aceptaría que hay heterocedasticidad.

$$e^2 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + U \quad (7)$$

$$H_0 \rightarrow \sigma^2 = \text{Constante} \rightarrow \text{No hay problema de Heterocedasticidad}$$

$$H_1 \rightarrow \sigma^2 \neq \text{Constante} \rightarrow \text{Hay problema de Heterocedasticidad}$$

Solucion de la Heterocedasticidad

Cuando solo hay una variable explicativa que produce heterocedasticidad el problema se puede solucionar de la siguiente forma: se divide el modelo inicial entre la variable que genera problemas de heterocedasticidad (X_p) elevada a C. Puesto que C no puede calcularse directamente, se calcula a partir de h (8).

$$\frac{Y}{X_p^C} = \frac{\beta_0}{X_p^C} + \beta_1 \frac{X_1}{X_p^C} + \beta_2 \frac{X_2}{X_p^C} + \dots + \beta_K \frac{X_K}{X_p^C} + U \quad C = \frac{h}{2} \quad (8)$$

Para calcular h , se procede a ajustar los Residuos² frente a la variable explicativa que genera el problema de heterocedasticidad (X_P) elevada a h (9).

$$\text{Residuos}^2 = \beta_0 + \beta_1 X_P^h \quad (9)$$

El valor de “ h ” puede oscilar entre valores de tipo Real, aunque se recomienda el uso de Enteros positivos, Enteros negativos, y fraccionarios entre 0 y 1. Como por ejemplo: 1, 2, 3,4, -1, -2, -3, $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$... Se realiza una tabla con las diferentes posibilidades y se escoge la h que de un ajuste del modelo (R^2) mayor. Se debe tener precaución al realizar el ajuste de la solución, y de que quitar la constante que pone por defecto el programa, ya que esta se incluye manualmente como $1/(X)^h$, para forzar a dividir el parámetro β_0 por la X del problema. Por último, se debe comprobar que el modelo solución no tiene heterocedasticidad (repetir el proceso de análisis para el modelo solución).

3.2.5. Autocorrelación

Una de las hipótesis del modelo hace referencia a que la perturbación en un momento dado no debe depender de los valores tomados anteriormente. Si ocurre esto, existe autocorrelación, y el modelo sufrirá los siguientes problemas:

- Las estimaciones de los parámetros no tendrán mínima varianza.
- Los valores “ t ” y “ F ” ya no servirían, ya que dejan de seguir la distribución correspondiente, y son muy grandes. Al hacerse muy grandes pueden aceptarse como variables explicativas aquellas que no lo son.
- Las predicciones que ofrece el modelo no son adecuadas.

Las causas del problema de autocorrelación son las siguientes:

- Las variables estudiadas son series temporales
- El modelo no ha sido formulado correctamente. Cuando se establece una relación incorrecta entre la variable a explicar y las variables explicativas, se produce un error de especificación del modelo.
- La variable a explicar se ha medido con error.

Estos problemas se pueden identificar mediante diferentes métodos gráficos y pruebas de autocorrelación.

Si se representan los residuos frente a la estimación de la variable, frente al índice o frente a las variables explicativas es posible apreciar si existe cierta correlación entre ellos o no. El gráfico de residuos frente a estimación y frente a variable explicativa

permite identificar si el origen de la autocorrelación es una variable explicativa en particular, mientras que el gráfico de los residuos frente al índice muestra los residuos según un orden temporal o espacial, siendo posiblemente éstos los causantes de la autocorrelación.

Es posible definir algunas pruebas para comprobar la existencia de autocorrelación, como las que se presentan a continuación:

1. Representación de la FAS y la FAP.
2. Prueba de Durbin-Watson.
3. Prueba de Wallis.
4. Prueba de Durbin.
5. Prueba de Ljung-Box.
6. Prueba de Breusch-Godfrey.

La primera prueba es la más general, y estudia la correlación de cualquier orden, además de medirla de acuerdo a dos criterios. La función de autocorrelación simple (FAS) es el conjunto de los coeficientes de autocorrelación simple ρ_s (con $s = 1, 2, 3, 4, \dots$), miden la relación lineal que hay entre el residuo en el instante de tiempo t , e_t , y el residuo en el instante $t-s$, e_{t-s} (Figura III. 12. A).

La función de autocorrelación parcial (FAP) es el conjunto de coeficientes de autocorrelación parcial α_s (con $s = 1, 2, 3, 4, \dots$), y miden la relación directa que existe entre el residuo en el instante de tiempo t , e_t , y el residuo en el instante $t-s$, e_{t-s} , eliminando el posible efecto de los residuos intermedios a ambos (Figura III. 12.B).

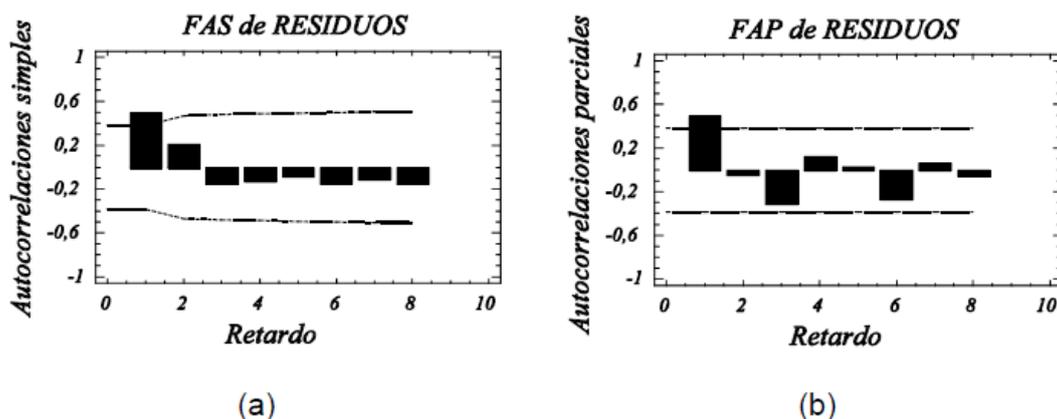


Figura III. 12. Funciones de autocorrelación simples y parciales de los residuos del ajuste.

Si algún coeficiente de autocorrelación poblacional, simple o parcial, es distinto de cero, esto implica que existe autocorrelación. La detección de tales coeficientes significativos se hace mediante una prueba de hipótesis, que aparece en los propios gráficos, en

forma de una banda simétrica respecto al eje de abscisas. Si algún coeficiente de correlación sobrepasa tal banda, eso significa dos cuestiones, la primera es que el correspondiente coeficiente de correlación poblacional es distinto de cero y que existe autocorrelación, y en segundo lugar indica su orden (para el retardo s implica la relación entre e_t y e_{t-s})(Figura III. 12). Si ninguno de los retardos supere la banda, no habría problemas de autocorrelación y en caso contrario se procedería a solucionar el problema. Existen cuatro métodos de resolución de problemas de autocorrelación. A continuación se van a nombrar estos métodos:

- Transformación de Cochrane-Orcutt.
- Método de las diferenciaciones.
- Método de Cochrane-Orcutt dos etapas.
- Método de Durbin dos etapas.

3.2.6. Puntos Atípicos e Influyentes

Puntos influyentes

Es posible encontrar modelos en los que la estimación de los parámetros este determinada por un par de puntos influyentes, en vez de por la mayoría de puntos. Para analizar la importancia de los puntos influyentes, tenemos los ensayos de robustez del modelo, y los del parámetro. En Statgraphics esta relación d/σ se llama “leverage o influencia” y se representa por “ n_{ii} ” (10).

$$\bar{n} = \frac{\sum n_{ii}}{n} \hat{\sigma}_n^2 = \frac{\sum n_{ii}^2}{n} - \bar{n}^2 \quad (10)$$

Existen dos criterios para saber si un punto es influyente:

- ✓ A priori : si el *leverage* queda fuera del intervalo $n_{ii} > 2\bar{n}$, tal que $\bar{n} \pm 3\hat{\sigma}_n$, se considera punto influyente
 $n_{ii} \geq 2 \cdot \text{influencia}$
 $\text{Influencia} \geq 2 \cdot \text{Influencia media de un solo punto}$
- ✓ A posteriori: donde el criterio a seguir para ver si el punto influye sobre los parámetros es que $|DFITS| \geq \sqrt[2]{\bar{n}}$.

Puntos atípicos

Básicamente se distinguen cuando generan residuos muy grandes tomándose dos límites para su identificación y la posibilidad de eliminación: si $|rt_i| \geq 2 \rightarrow$ es un punto anómalo, y si $|rt_i| \geq 3 \rightarrow$ es un punto anómalo candidato a eliminar del modelo.

3.3. Predicción

Uno de los objetivos deseados al realizar modelos económicos es el de poder hacer predicciones de los valores de la variable, si bien en realidad lo que permite el modelo ajustado es predecir el valor medio de dicha variable, o encontrar un intervalo que contenga con una probabilidad determinada al valor real. Existen dos formas de realizar predicciones: predicción puntual y predicción por intervalos de confianza. Ambas, están representadas en la Figura III. 13 [14].

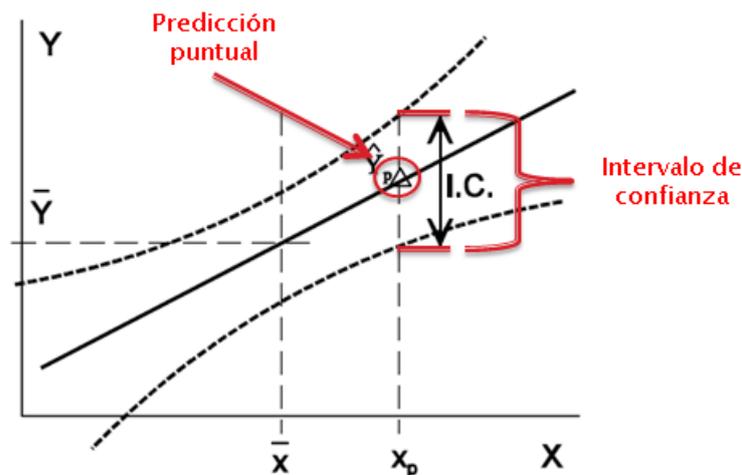


Figura III. 13 Grafico predicción puntual y predicción por intervalos de confianza.

4. Herramientas de cálculo

En este apartado se comentan los programas empleados para realizar todos los cálculos requeridos en este trabajo. Fundamentalmente se han necesitado los siguientes:

- Excel: herramienta utilizada para la realización de tablas y gráficos, así como alguno de los cálculos matemáticos requeridos, aprovechando la capacidad de cálculo del software.
- Statgraphics: programa utilizado en la versión centurión XVI para la representación de cada una de las regresiones múltiples lineales planteadas.

IV. Cálculos y Resultados

1. Matriz de Datos

En primer lugar, tal y como se indica en el apartado Búsqueda de Metodología y Técnicas se procede a realizar una búsqueda de las variables (por comunidades autónomas) que pueden afectar al flujo migratorio. En las siguientes tablas (Tabla IV. 1 y Tabla IV. 2) se encuentra la matriz de datos inicial con la que se empieza a trabajar en este modelo:

Tabla IV. 1. Matriz de datos inicial (parte A)

| CCAA | Flujo_Migración | BECAS_CCAA | | | BECAS_MICIN | | |
|----------------------|-----------------|---------------------|-------------------------|----------------|----------------------|--------------------------|----------------------|
| | | Becas_CCAA OBLIG | Becas_CCAA PostNOUNI | Becas_ Univ | Becas_MICIN OBLIG | Becas_MICIN PostNOUNI | Becas_MICI N_Univ |
| Andalucía | 50575 | 85804.3 | 50977.7 | 322.1 | 13620.2 | 140627.7 | 266148 |
| Aragón | 10771 | 6463.2 | 4.5 | 680.4 | 1172.6 | 8105.5 | 19696.5 |
| Asturias | 5119 | 1691 | 240.4 | 468.2 | 778.6 | 8122.2 | 16655.9 |
| Balears, Illes | 13866 | 589.5 | 0 | 0.9 | 2759.4 | 4083.7 | 8694.2 |
| Canarias | 16606 | 8487.9 | 28.1 | 7528.7 | 4810 | 27175.2 | 39296.8 |
| Cantabria | 3267 | 1397.1 | 285.1 | 580 | 540.4 | 5255.7 | 7297 |
| Castilla y León | 12773 | 11730.1 | 510 | 4941.5 | 3509.1 | 22479.7 | 70209.7 |
| Castilla - La Mancha | 17294 | 283.8 | 225.4 | 4811.9 | 1441.4 | 26008.5 | 31845.1 |
| Cataluña | 139014 | 356.4 | 30.6 | 3102.7 | 12418.6 | 38389.5 | 118968.2 |
| Comunitat Valenciana | 71087 | 58089.1 | 7.8 | 13619.1 | 9256.5 | 57925.2 | 122645 |
| Extremadura | 3404 | 54.4 | 101.6 | 776 | 700 | 23787 | 36539.5 |
| Galicia | 12300 | 19533.9 | 852.5 | 1761.4 | 6819.4 | 29870.9 | 63945.3 |
| Madrid | 125868 | 46001.1 | 142 | 8646.4 | 8093.3 | 33462.2 | 135056.4 |
| Murcia | 14896 | 5295.3 | 16.1 | 223.2 | 4759.9 | 18599.4 | 41212.9 |
| Navarra | 6463 | 130.4 | 389.5 | 3586.4 | 595 | 2994.1 | 6693.7 |
| País Vasco | 21158 | 44355.7 | 6930.7 | 19693.8 | 3.2 | 388.2 | 7739.2 |
| Rioja, La | 4438 | 784 | 101.5 | 385.1 | 464.5 | 1749.7 | 3621.1 |
| Ceuta | 1188 | 0 | 0 | 0 | 448.1 | 955.6 | 0 |
| Melilla | 2215 | 0 | 0 | 0 | 317.5 | 1528.5 | 0 |

Tabla IV. 2. Matriz de datos inicial (parte B)

| CCAA | PIB | IPC | IPI | DESEMP_ ESO | I+D+i_ PUB | I+D+i_Empresa | PARO | PARO_JUVENIL | IPVIVIENDA |
|----------------------|-------|----------|----------|----------------|---------------|---------------|--------|--------------|------------|
| Andalucía | 16843 | 1241.533 | 1029.469 | 80.47 | 304689 | 538425 | 144.89 | 1626.9 | 293.867 |
| Aragón | 24693 | 1244.447 | 1124.776 | 81.66 | 67679 | 162051 | 85.53 | 145 | 229.692 |
| Asturias | 20035 | 1244.072 | 1085.389 | 83.46 | 27810 | 96662 | 96.51 | 147.9 | 279.993 |
| Balears, Illes | 23625 | 1249.124 | 1068.873 | 72.68 | 32898 | 11120 | 89.34 | 170.7 | 269.337 |
| Canarias | 19312 | 1233.337 | 1118.341 | 83.19 | 62369 | 43201 | 134.92 | 423.8 | 285.379 |
| Cantabria | 20661 | 1257.26 | 1134.72 | 82.06 | 20481 | 38701 | 81.79 | 70 | 251.252 |
| Castilla y León | 21395 | 1252.388 | 1042.258 | 84.46 | 48868 | 298224 | 86.99 | 282.6 | 270.243 |
| Castilla - La Mancha | 18273 | 1244.816 | 1049.125 | 84.24 | 25416 | 117670 | 119.87 | 340.6 | 274.947 |
| Cataluña | 26512 | 1255.926 | 1143.429 | 76.74 | 596591 | 1675742 | 92.48 | 887.9 | 218.122 |
| Comunitat Valenciana | 19695 | 1245.055 | 1117.815 | 76.86 | 122972 | 405295 | 112.2 | 709.5 | 272.283 |
| Extremadura | 15497 | 1244.275 | 978.458 | 86.27 | 29239 | 26886 | 135.47 | 194.5 | 306.517 |
| Galicia | 19893 | 1246.666 | 1078.681 | 87.3 | 69672 | 214937 | 88.17 | 329.2 | 294.087 |
| Madrid | 30678 | 1244.357 | 998.144 | 79.55 | 860871 | 1946000 | 79.05 | 669.6 | 232.81 |
| Murcia | 18392 | 1248.333 | 1039.399 | 80.48 | 34397 | 87291 | 115.92 | 252.8 | 289.716 |
| Navarra | 27795 | 1245.985 | 1070.571 | 83.91 | 25065 | 217831 | 71.72 | 74.7 | 231.751 |
| País Vasco | 29313 | 1247.084 | 1084.182 | 87.41 | 92379 | 995730 | 66.29 | 193.2 | 244.461 |
| Rioja | 24426 | 1246.318 | 1045.007 | 80.22 | 14993 | 30283 | 80.14 | 28.7 | 240.224 |
| Ceuta | 18455 | 1233.226 | 0 | 85.24 | 0 | 0 | 139.16 | 17.8 | 278.949 |
| Meiilla | 16836 | 1221.314 | 0 | 80.52 | 0 | 0 | 129.63 | 17.7 | 295.701 |

1.1. Análisis Bivariante (Multicolinealidad)

En este apartado se va a proceder a realizar las tres pruebas para comprobar que no existe problema de multicolinealidad, y en el caso de que exista, solucionar el problema, tal y como se ha indicado en el apartado de Análisis Bivariante.

1.1.1. Identificación

En primer lugar como se observa en la matriz de correlación inicial (Tabla IV. 1 y Tabla IV. 2), todos los valores mayores que $|0.7|$ indican que existe un problema de correlación por lo que se procederá a la eliminación de variables.

| | Becas_CCAA_OBLIG | Becas_CCAA_PostNOUNI | Becas_Univ | Becas_MICIN_OBLIG | Becas_MICIN_PostNOUNI | Becas_MICIN_Univ |
|-----------------------|------------------|----------------------|------------|-------------------|-----------------------|------------------|
| Becas_CCAA_OBLIG | 1,000 | 0,727 | 0,462 | 0,618 | 0,774 | 0,793 |
| Becas_CCAA_PostNOUNI | 0,727 | 1,000 | -0,086 | 0,520 | 0,855 | 0,753 |
| Becas_Univ | 0,462 | -0,086 | 1,000 | 0,044 | -0,030 | 0,045 |
| Becas_MICIN_OBLIG | 0,618 | 0,520 | 0,044 | 1,000 | 0,812 | 0,903 |
| Becas_MICIN_PostNOUNI | 0,774 | 0,855 | -0,030 | 0,812 | 1,000 | 0,948 |
| Becas_MICIN_Univ | 0,793 | 0,753 | 0,045 | 0,903 | 0,948 | 1,000 |
| PIB | 0,012 | -0,255 | 0,405 | -0,079 | -0,385 | -0,160 |
| IPC | -0,278 | -0,230 | -0,164 | -0,031 | -0,262 | -0,138 |
| IPI | -0,162 | -0,219 | 0,141 | 0,059 | -0,178 | -0,191 |
| DESEMP_ESO | -0,076 | -0,029 | 0,180 | -0,421 | -0,173 | -0,256 |
| I+D+i_PUB | 0,401 | 0,166 | 0,186 | 0,655 | 0,361 | 0,609 |
| I+D+i_Empres | 0,399 | 0,091 | 0,432 | 0,544 | 0,222 | 0,482 |
| PARO | 0,233 | 0,446 | -0,208 | 0,376 | 0,635 | 0,459 |
| PARO_JUVENIL | 0,758 | 0,779 | 0,075 | 0,909 | 0,950 | 0,971 |
| IPVIVIENDA | 0,139 | 0,266 | -0,199 | 0,065 | 0,349 | 0,160 |

Figura IV. 1 Matriz de correlación (parte A)

| | PIB | IPC | IPI | DESEMP_ESO | I+D+i_PUB | I+D+i_Empresa | PARO | PARO_JUVENIL | IPVIVIENDA |
|-----------------------|--------|--------|--------|------------|-----------|---------------|--------|--------------|------------|
| Becas_CCAA_OBLIG | 0,012 | -0,278 | -0,162 | -0,076 | 0,401 | 0,399 | 0,233 | 0,758 | 0,139 |
| Becas_CCAA_PostNOUNI | -0,255 | -0,230 | -0,219 | -0,029 | 0,166 | 0,091 | 0,446 | 0,779 | 0,266 |
| Becas_Univ | 0,405 | -0,164 | 0,141 | 0,180 | 0,186 | 0,432 | -0,208 | 0,075 | -0,199 |
| Becas_MICIN_OBLIG | -0,079 | -0,031 | 0,059 | -0,421 | 0,655 | 0,544 | 0,376 | 0,909 | 0,065 |
| Becas_MICIN_PostNOUNI | -0,385 | -0,262 | -0,178 | -0,173 | 0,361 | 0,222 | 0,635 | 0,950 | 0,349 |
| Becas_MICIN_Univ | -0,160 | -0,138 | -0,191 | -0,256 | 0,609 | 0,482 | 0,459 | 0,971 | 0,160 |
| PIB | 1,000 | 0,228 | 0,158 | -0,157 | 0,485 | 0,640 | -0,813 | -0,162 | -0,870 |
| IPC | 0,228 | 1,000 | 0,224 | -0,204 | 0,078 | 0,189 | -0,519 | -0,176 | -0,389 |
| IPI | 0,158 | 0,224 | 1,000 | -0,204 | -0,128 | -0,028 | -0,249 | -0,075 | -0,387 |
| DESEMP_ESO | -0,157 | -0,204 | -0,204 | 1,000 | -0,336 | -0,199 | -0,019 | -0,293 | 0,268 |
| I+D+i_PUB | 0,485 | 0,078 | -0,128 | -0,336 | 1,000 | 0,931 | -0,087 | 0,582 | -0,418 |
| I+D+i_Empresa | 0,640 | 0,189 | -0,028 | -0,199 | 0,931 | 1,000 | -0,272 | 0,476 | -0,533 |
| PARO | -0,813 | -0,519 | -0,249 | -0,019 | -0,087 | -0,272 | 1,000 | 0,512 | 0,734 |
| PARO_JUVENIL | -0,162 | -0,176 | -0,075 | -0,293 | 0,582 | 0,476 | 0,512 | 1,000 | 0,138 |
| IPVIVIENDA | -0,870 | -0,389 | -0,387 | 0,268 | -0,418 | -0,533 | 0,734 | 0,138 | 1,000 |

Figura IV. 2 Matriz de Correlación (parte B)

Para el criterio de eliminación de variables se calcula la matriz de correlación inversa, dónde el valor de la diagonal indicará el orden en el que se excluirán las variables que presenten problemas de correlación.

| INVERSA | Becas_CCAA_O BLIG | Becas_CCAA_ PostNOUNI | Becas_Univ | Becas_MICIN_ OBLIG | Becas_MICIN_ PostNOUNI | Becas_MICIN_ Univ | PIB |
|---------------------------|----------------------|--------------------------|------------|-----------------------|---------------------------|----------------------|----------|
| Becas_CCAA_O BLIG | 331.022 | -253.187 | -153.826 | -78.934 | -779.619 | -93.772 | 118.490 |
| Becas_CCAA_P ostNOUNI | -253.187 | 229.354 | 131.215 | 92.154 | 540.230 | 80.092 | -128.586 |
| Becas_Univ | -153.826 | 131.215 | 84.919 | 51.848 | 315.869 | 40.605 | -74.835 |
| Becas_MICIN_O BLIG | -78.934 | 92.154 | 51.848 | 66.067 | 156.319 | 0.349 | -66.061 |
| Becas_MICIN_P ostNOUNI | -779.619 | 540.230 | 315.869 | 156.319 | 2298.484 | 74.938 | -178.906 |
| Becas_MICIN_U niv | -93.772 | 80.092 | 40.605 | 0.349 | 74.938 | 175.464 | -20.490 |
| PIB | 118.490 | -128.586 | -74.835 | -66.061 | -178.906 | -20.490 | 146.781 |
| IPC | 89.117 | -73.546 | -40.542 | -25.863 | -201.522 | -25.599 | 54.267 |
| IPI | -65.631 | 43.861 | 23.081 | 4.105 | 189.477 | 28.027 | -2.106 |
| DESEMP_ESO | 129.114 | -99.229 | -56.261 | -29.100 | -327.580 | -34.029 | 50.016 |
| I+D+i_PUB | 9.251 | 10.458 | 21.460 | 18.741 | -89.481 | -29.167 | -14.092 |
| I+D+i_Empresa | -350.404 | 254.598 | 126.658 | 66.091 | 1004.685 | 88.651 | -114.576 |
| PARO | 99.591 | -81.961 | -54.078 | -32.207 | -192.653 | -4.194 | 85.965 |
| PARO_ JUVENIL | 1045.125 | -817.621 | -458.605 | -271.925 | -2746.659 | -277.043 | 328.068 |
| IPVIVIENDA | -39.584 | 9.133 | 5.886 | -16.044 | 149.043 | 14.102 | 32.137 |

Figura IV. 3 Matriz de Correlación Inversa (parte A)

| INVERSA | IPC | IPI | DESEMP_E SO | I+D+i_PUB | I+D+i_ Empresa | PARO | PARO_ JUVENIL | IPVIVIENDA |
|---------------------------|----------|----------|----------------|-----------|-------------------|----------|------------------|------------|
| Becas_CCAA_O BLIG | 89.117 | -65.631 | 129.114 | 9.251 | -350.404 | 99.591 | 1045.125 | -39.584 |
| Becas_CCAA_P ostNOUNI | -73.546 | 43.861 | -99.229 | 10.458 | 254.598 | -81.961 | -817.621 | 9.133 |
| Becas_Univ | -40.542 | 23.081 | -56.261 | 21.460 | 126.658 | -54.078 | -458.605 | 5.886 |
| Becas_MICIN_O BLIG | -25.863 | 4.105 | -29.100 | 18.741 | 66.091 | -32.207 | -271.925 | -16.044 |
| Becas_MICIN_P ostNOUNI | -201.522 | 189.477 | -327.580 | -89.481 | 1004.685 | -192.653 | -2746.659 | 149.043 |
| Becas_MICIN_U niv | -25.599 | 28.027 | -34.029 | -29.167 | 88.651 | -4.194 | -277.043 | 14.102 |
| PIB | 54.267 | -2.106 | 50.016 | -14.092 | -114.576 | 85.965 | 328.068 | 32.137 |
| IPC | 33.151 | -15.090 | 38.447 | 12.309 | -111.486 | 38.629 | 283.897 | -2.630 |
| IPI | -15.090 | 22.553 | -27.678 | -12.691 | 88.934 | -7.151 | -240.621 | 20.186 |
| DESEMP_ESO | 38.447 | -27.678 | 55.844 | 15.889 | -160.905 | 39.863 | 432.772 | -16.417 |
| I+D+i_PUB | 12.309 | -12.691 | 15.889 | 76.598 | -107.605 | -8.821 | 96.222 | -13.896 |
| I+D+i_Empresa | -111.486 | 88.934 | -160.905 | -107.605 | 540.451 | -92.494 | -1277.097 | 58.837 |
| PARO | 38.629 | -7.151 | 39.863 | -8.821 | -92.494 | 68.521 | 254.054 | 5.231 |
| PARO_ JUVENIL | 283.897 | -240.621 | 432.772 | 96.222 | -1277.097 | 254.054 | 3659.745 | -143.932 |
| IPVIVIENDA | -2.630 | 20.186 | -16.417 | -13.896 | 58.837 | 5.231 | -143.932 | 32.762 |

Figura IV. 4 Matriz de Correlación Inversa (parte B)

Como se puede observar en las tablas de la Figura IV. 3 y Figura IV. 4, se empezara eliminando las variables con mayor resultado en la diagonal, en este caso, se prescindirá de la variable PARO_JUVENIL, y así sucesivamente hasta que ninguno de los valores de la diagonal sea mayor que 10.

1.1.2. Solución y comprobación

Tras eliminar todas las variables en las que se observaron problemas de correlación quedó la matriz de correlación representada en la Figura IV. 5, dónde se observa que ya no existen problemas de multicolinealidad por la correlación entre pares de variables.

| | Becas_CCAA_PostNOUNI | Becas_Univ | Becas_MICIN_OBLIG | IPC | IPI | DESEMP_ESO | I+D+i_PUB | PARO |
|----------------------|----------------------|------------|-------------------|-----------|-----------|------------|------------|----------|
| Becas_CCAA_PostNOUNI | 1 | -0.0860049 | 0.519695 | -0.229917 | -0.219371 | -0.0289942 | 0.166241 | 0.445573 |
| Becas_Univ | -0.0860049 | 1 | 0.0436277 | -0.16419 | 0.140907 | 0.179524 | 0.186071 | -0.20849 |
| Becas_MICIN_OBLIG | 0.519695 | 0.0436277 | 1 | -0.030863 | 0.0586468 | -0.420611 | 0.654665 | 0.375975 |
| IPC | -0.229917 | -0.16419 | -0.0308631 | 1 | 0.223743 | -0.203586 | 0.0777675 | -0.519 |
| IPI | -0.219371 | 0.140907 | 0.0586468 | 0.223743 | 1 | -0.203693 | -0.128378 | -0.24889 |
| DESEMP_ESO | -0.0289942 | 0.179524 | -0.420611 | -0.203586 | -0.203693 | 1 | -0.336078 | -0.01879 |
| I+D+i_PUB | 0.166241 | 0.186071 | 0.654665 | 0.0777675 | -0.128378 | -0.336078 | 1 | -0.08672 |
| PARO | 0.445573 | -0.208491 | 0.375975 | -0.519001 | -0.248885 | -0.0187859 | -0.0867227 | 1 |

Figura IV. 5 Matriz de Correlación Final

Se procede a realizar la matriz inversa de correlación, tal y como se explica en el apartado Matriz de correlación Inversa. En la Figura IV. 6 se puede comprobar que ninguno de los valores de la diagonal son mayores que 10, por tanto, no existe problema de multicolinealidad del tipo una variable con todas las demás.

| INVERSA | Becas_CCAA_PostNOUNI | Becas_Univ | Becas_MICIN_OBLIG | IPC | IPI | DESEMP_ESO | I+D+i_PUB | PARO |
|----------------------|----------------------|-------------|-------------------|-----------|------------|--------------|------------|----------|
| Becas_CCAA_PostNOUNI | 1.803783422 | 0.137919811 | -1.503476746 | 0.224186 | 0.45920527 | -0.255864861 | 0.62034032 | 0.069941 |
| Becas_Univ | 0.137919811 | 1.359100787 | -0.303417937 | 0.504357 | -0.2360329 | -0.380790126 | -0.2302879 | 0.511877 |
| Becas_MICIN_OBLIG | -1.503476746 | -0.30341794 | 4.790075287 | -0.72257 | -1.1730922 | 0.627062505 | -2.8953316 | -2.10059 |
| IPC | 0.224185982 | 0.50435697 | -0.722570026 | 1.7336981 | -0.0009232 | 0.104043441 | 0.34675499 | 1.208518 |
| IPI | 0.459205273 | -0.23603285 | -1.173092188 | -0.000923 | 1.56159949 | 0.255993783 | 1.08058853 | 0.673933 |
| DESEMP_ESO | -0.255864861 | -0.38079013 | 0.627062505 | 0.1040434 | 0.25599378 | 1.470850507 | 0.2187727 | -0.03683 |
| I+D+i_PUB | 0.620340317 | -0.23028794 | -2.895331611 | 0.346755 | 1.08058853 | 0.218772696 | 3.14972325 | 1.490323 |
| PARO | 0.069940507 | 0.511876556 | -2.100590478 | 1.2085183 | 0.67393331 | -0.036829194 | 1.49032326 | 2.788835 |

Figura IV. 6. Matriz de Correlación Inversa Final

Finalmente, se realizará el tercero de los métodos para comprobar que no existen problemas de Multicolinealidad, que es el cálculo del índice de acondicionamiento (11).

$$I.C = \sqrt{\frac{AUTOVALOR\ MAX}{AUTOVALOR\ MIN}} = \sqrt{\frac{2,33093}{0,10897}} = 4,62699 \quad (11)$$

Como se refleja en el índice de acondicionamiento, no existe multicolinealidad puesto que el valor calculado es inferior a 10. Una vez realizadas las tres pruebas y comprobado que no hay problemas de multicolinealidad, se procederá con el planteamiento del modelo de regresión múltiple lineal.

2. Regresión Múltiple Lineal

En este trabajo se decide intentar modelizar el comportamiento del flujo de migración mediante la técnica de regresión múltiple lineal. Esta técnica consiste básicamente en explicar la influencia de dos o más variables independientes "X" en una única variable que denominaremos "Y" dependiente.

2.1. Modelo teórico

En este proyecto la variable dependiente será el "Flujo de emigración de los españoles" (Y) y la unidad de medida utilizada para esta variable será número de movimientos migratorios en personas. Para el planteamiento del modelo teórico (12) se utilizarán las siguientes variables explicativas, libres de multicolinealidad:

- X1 = Becas postobligatorias No Universitarias por CCAA (Becas_CCAA).
- X2 = Becas Universitarias por CCAA (Becas_Univ).
- X3 = Becas enseñanza obligatoria del Ministerio de Educación (Becas_MICIN).
- X4 = Índice de precios de Consumo por CCAA (IPC).
- X5 = Índice de producción industrial por CCAA (IPI).
- X6 = Tasa de Desempleo al finalizar la enseñanza obligatoria (Desempleo_ESO).
- X7 = Inversión en I+D+I público por CCAA (I+D+I).
- X8 = Tasa de Paro por CCAA (PARO)

$$\text{FLUJO_EMIGRACIÓN} = \beta_0 + \beta_1 * \text{Becas_CCAA} + \beta_2 * \text{Becas_Univ} + \beta_3 * \text{Becas_MICIN} + \beta_4 * \text{IPC} + \beta_5 * \text{IPI} + \beta_6 * \text{DESEMP_ESO} + \beta_7 * \text{I+D+I} + \beta_8 * \text{PARO} + U \quad (12)$$

2.2. Estimación de la primera ecuación de Regresión Múltiple y validación del modelo.

En este apartado se han estimado los parámetros del modelo utilizando el programa Statgraphics. Se procederá a validar paso a paso el modelo tal y como se explica detalladamente en el apartado 3.2 de Metodología. A continuación se contempla la función estimada del modelo que proporciona el programa (13).

$$\text{FLUJO_EMIGRACIÓN} = -1.28079\text{E}6 - 0.494948 * \text{Becas_CCAA} + 1.40154 * \text{Becas_Univ} + +2.41433 * \text{Becas_MICIN} + 1026.08 * \text{IPC} + 69.6088 * \text{IPI} - 1200.41 * \text{DESEMP_ESO} + +0.129853 * \text{I+D+I} + 238.889 * \text{PARO} \quad (13)$$

2.2.1. Análisis de la Significación

En este punto se procederá a identificar si las variables y el modelo son significativos. En la Tabla IV. 3 se observan los P.Valores de todas las variables explicativas. En este caso, de forma general, la mayoría de los P.Valores son mayores que 0'05. Esto indica que los parámetros del modelo no son significativos. Exceptuando las variables Becas_Univ y I+D+I que son menores que 0'05, por tanto, si que son significativas. Ocurre lo mismo en el caso de la constante, que no es significativa.

Tabla IV. 3. Resumen análisis regresión múltiple.

| | | <i>Error</i> | <i>Estadístico</i> | |
|------------------|-------------------|-----------------|--------------------|----------------|
| <i>Parámetro</i> | <i>Estimación</i> | <i>Estándar</i> | <i>T</i> | <i>Valor-P</i> |
| CONSTANTE | -1.28079E6 | 710546. | -1.80254 | 0.1091 |
| Becas_CCAA | -0.494948 | 0.249532 | -1.9835 | 0.0826 |
| Becas_Univ | 1.40154 | 0.486248 | 2.88235 | 0.0204 |
| Becas_MICIN | 2.41433 | 1.14638 | 2.10604 | 0.0683 |
| IPC | 1026.08 | 554.704 | 1.84979 | 0.1015 |
| IPI | 69.6088 | 60.6206 | 1.14827 | 0.2840 |
| DESEMP_ESO | -1200.41 | 705.101 | -1.70247 | 0.1271 |
| I+D+I | 0.129853 | 0.0172713 | 7.5184 | 0.0001 |
| PARO | 238.889 | 160.602 | 1.48746 | 0.1752 |

Tabla IV. 4. Resumen de análisis para el modelo de regresión.

| Fuente | Suma de Cuadrados | Gl | Cuadrado Medio | Razón-F | Valor-P |
|--|-------------------|----|----------------|---------|---------|
| Modelo | 2.76298E10 | 8 | 3.45372E9 | 41.17 | 0.0000 |
| Residuo | 6.71168E8 | 8 | 8.3896E7 | | |
| Total (Corr.) | 2.83009E10 | 16 | | | |
| R-cuadrada = 97.6285 por ciento | | | | | |
| R-cuadrado (ajustado para g.l.) = 95.2569 por ciento | | | | | |
| Error estándar del est. = 9159.47 | | | | | |
| Error absoluto medio = 5552.62 | | | | | |
| Estadístico Durbin-Watson = 1.11676 (P=0.0299) | | | | | |
| Autocorrelación de residuos en retraso 1 = 0.402448 | | | | | |

En cambio en la Tabla IV. 4 se puede ver un resumen del análisis para el modelo de regresión lineal, en el cual, se observa que el P.Valor del modelo es 0,000. Lo que indica que el modelo en sí es significativo.

2.2.2. Análisis de los gráficos de residuos.

A continuación se muestran los gráficos donde se comparan los residuos frente a cada una de variables explicativas con el fin de ir contemplando la posibilidad de la existencia de algún tipo de problema.

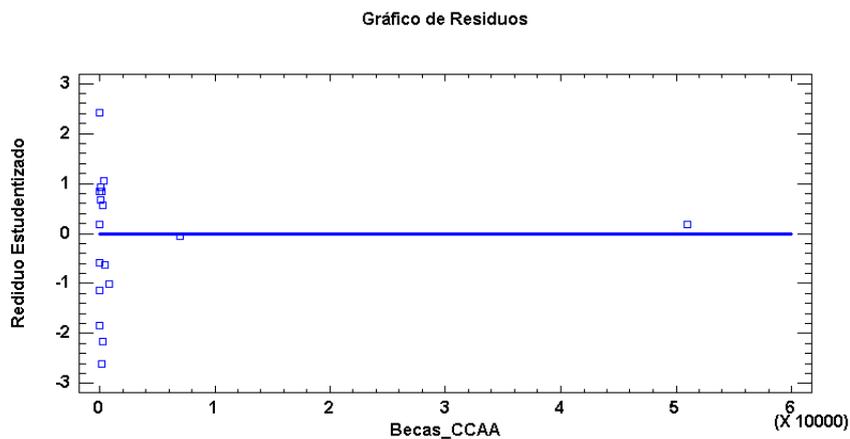


Figura IV. 7 Gráfico de Residuos frente a Becas_CCAA

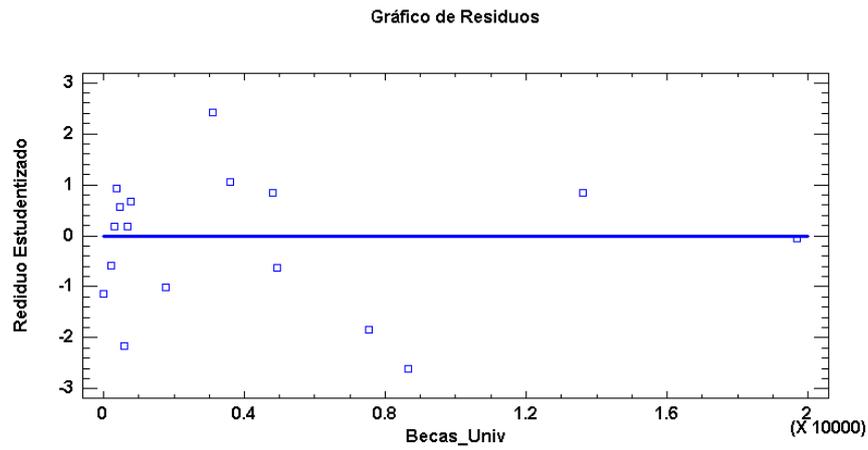


Figura IV. 8 Gráfico de Residuos frente Becas_Univ

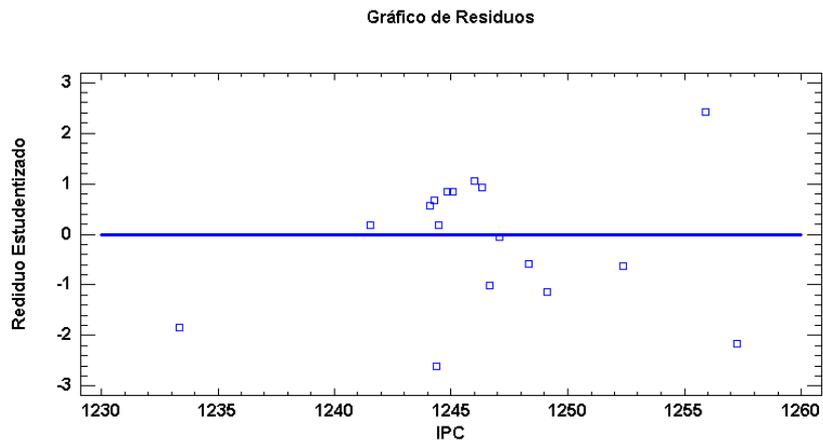


Figura IV. 9 Gráfico de Residuos frente IPC

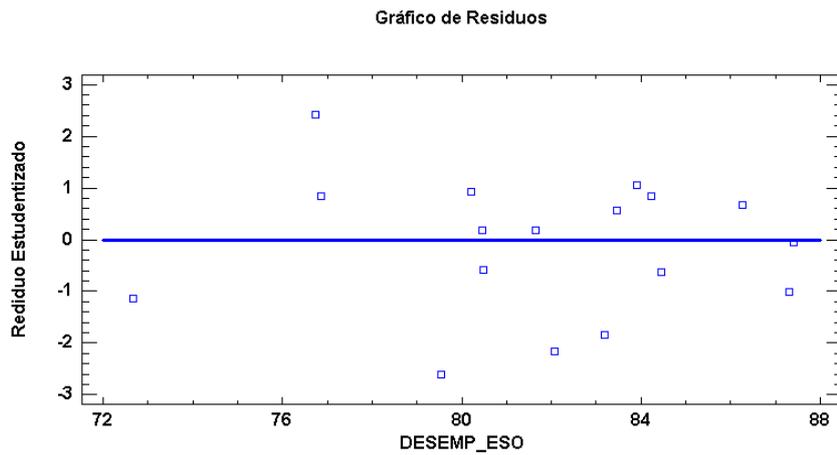


Figura IV. 10 Gráfico de Residuos frente DESEMP_ESO

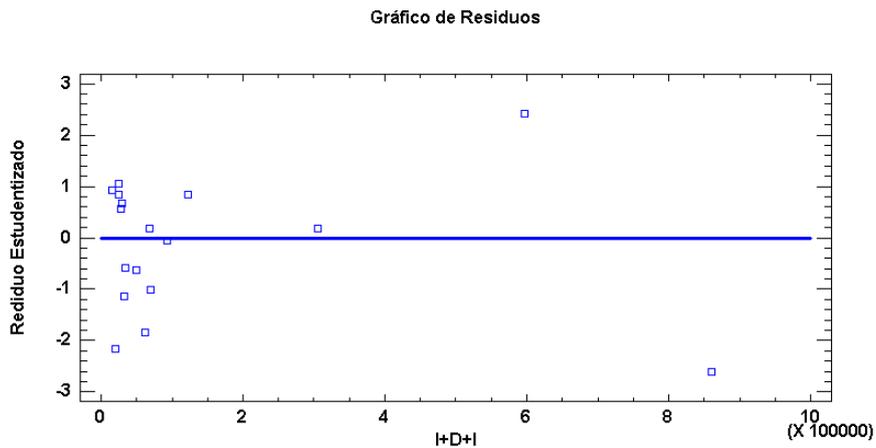


Figura IV. 11 Gráfico de Residuos frente I+D+I

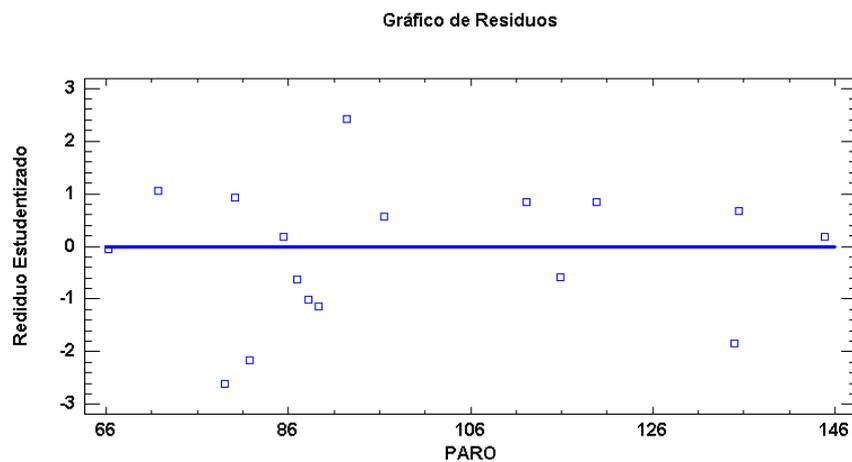


Figura IV. 12 Gráfico de Residuos frente a PARO

Como se puede ver en general de la Figura IV. 7 a la Figura IV. 12, no se observan formas en los residuos que se puedan asociar a ningún tipo de problemas. Lo único que se pueden apreciar son algunos residuos atípicos. Pero hay que destacar que en la Figura IV. 13 se intuye que podría haber problemas de falta de linealidad, ya que se observa una pequeña forma de parábola. Más adelante se solucionará este problema.

Por otro lado, En la Figura IV. 14 se observa un posible problema de heterocedasticidad debido a que los puntos presentan forma cónica, y también se ven algunos residuos atípicos. En apartados posteriores, se realizarán las pruebas pertinentes para comprobar si existe verdaderamente este problema.

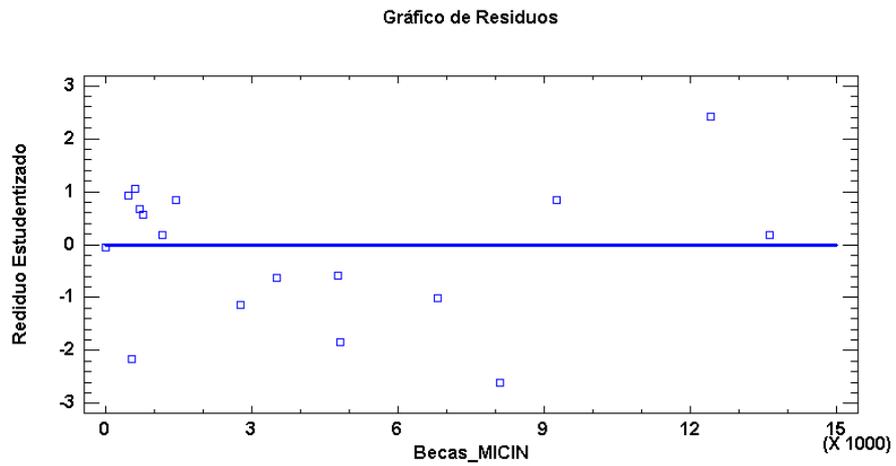


Figura IV. 13 Gráfico de residuos frente a la variable Becas_MICIN

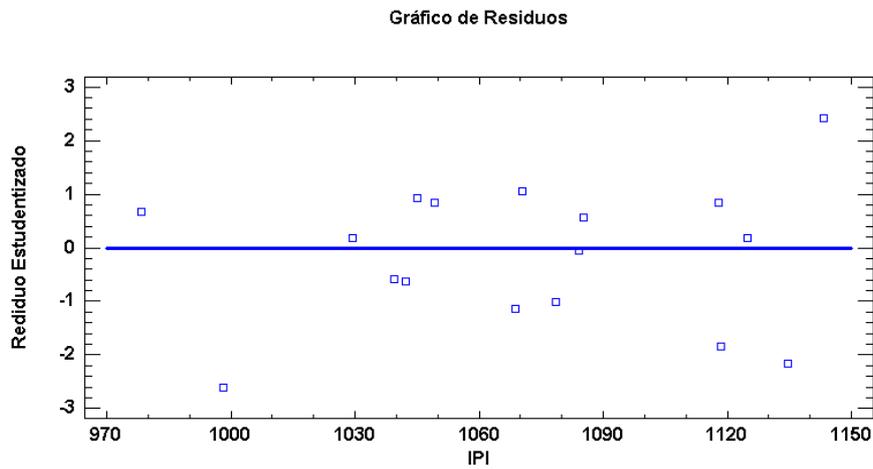


Figura IV. 14 Gráfico de residuos frente a la variable IPI

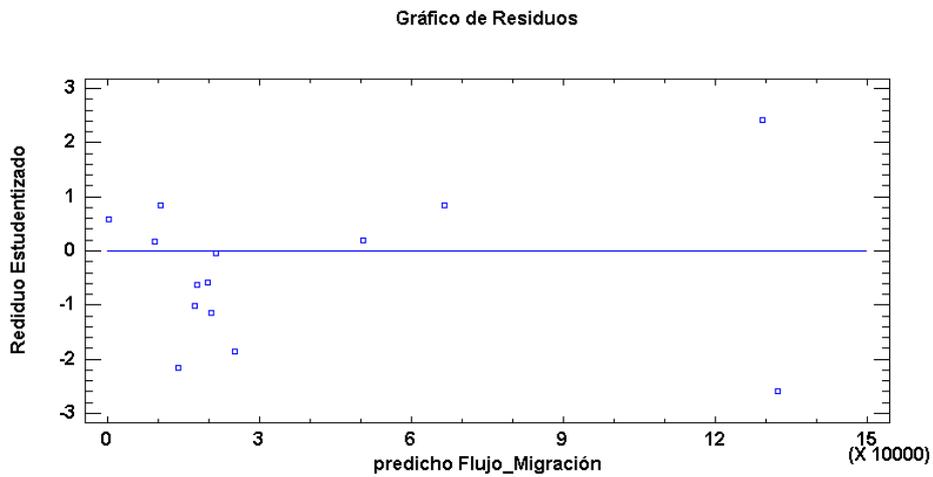


Figura IV. 15 Gráfico de residuos frente a la variable Flujo de Migración.

En la Figura IV. 15 Se detecta que la distribución de los residuos es aleatoria, aunque se puede intuir la forma parabólica a causa de la probable de falta de linealidad de una de las variables explicativas. Como se observa en la Figura IV. 16 los residuos se ordenan aleatoriamente por lo que no cabría esperar que pueda haber autocorrelación de primer orden.

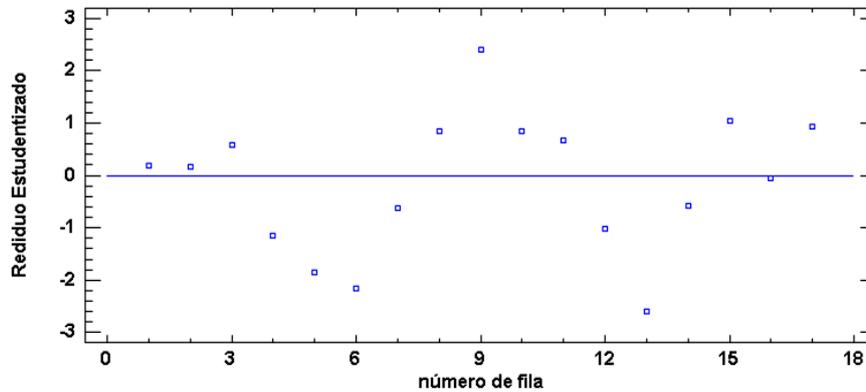


Figura IV. 16 Gráfico de residuos frente a número de fila.

SOLUCION AL PROBLEMA DE FALTA LINALIDAD

El problema de falta de linealidad, solo se puede observar de forma gráfica, por lo tanto, se procede a su corrección sin la necesidad de más análisis. Para la solución se eleva la variable *Becas_MICIN* al cuadrado de forma sucesiva para observar si se puede solucionar el problema. El procedimiento se repite hasta que deja de observarse la forma parabólica y el ajuste del modelo (R^2) deja de aumentar. En este caso, se ha elevado a cuatro la variable, porque con ocho dejaba de aumentar el ajuste del modelo. Como se observa en la Figura IV. 17 se ha solucionado el problema de falta de linealidad puesto que ya no se contempla forma de parábola.

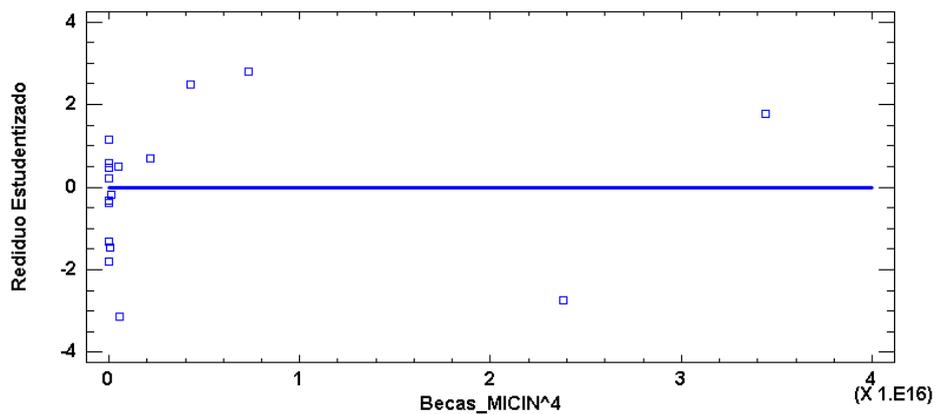


Figura IV. 17 Gráfico de residuos frente a la variable *Becas_MICIN*⁴

En la Tabla IV. 5 se observa la estimación de los parámetros del modelo al finalizar la corrección de la falta de linealidad. Se puede ver que el P.Valor de la mayoría de los parámetros ha cambiado, de forma que la mayoría de ellos son menores de 0'05, lo que indica que todas las variables que se encuentran por debajo de 0'05 son significativas para el modelo. En la

Tabla IV. 6 se comprueba que el modelo en sí sigue siendo significativo.

Tabla IV. 5. Resumen análisis regresión múltiple elevando la variable Becas_MICIN^4.

| | | <i>Error</i> | <i>Estadístico</i> | |
|------------------|-------------------|-----------------|--------------------|----------------|
| <i>Parámetro</i> | <i>Estimación</i> | <i>Estándar</i> | <i>T</i> | <i>Valor-P</i> |
| CONSTANTE | -126428. | 408216. | -0.309708 | 0.7647 |
| Becas_CCAA | -1.64701 | 0.237997 | -6.92027 | 0.0001 |
| Becas_Univ | 1.71096 | 0.237358 | 7.20835 | 0.0001 |
| Becas_MICIN^4 | 2.65209E-12 | 0 | 6.6474 | 0.0002 |
| IPC | 198.89 | 311.613 | 0.638261 | 0.5411 |
| IPI | -24.0771 | 34.8317 | -0.691242 | 0.5090 |
| DESEMP_ESO | -1256.96 | 337.363 | -3.72584 | 0.0058 |
| I+D+I | 0.108619 | 0.00919143 | 11.8174 | 0.0000 |
| PARO | 101.036 | 81.3267 | 1.24235 | 0.2493 |

Tabla IV. 6. Resumen análisis regresión del modelo elevando la variable Becas_MICIN^4.

| <i>Fuente</i> | <i>Suma de Cuadrados</i> | <i>Gl</i> | <i>Cuadrado Medio</i> | <i>Razón-F</i> | <i>Valor-P</i> |
|--|--------------------------|-----------|-----------------------|----------------|----------------|
| Modelo | 2.8141E10 | 8 | 3.51763E9 | 175.96 | 0.0000 |
| Residuo | 1.59927E8 | 8 | 1.99908E7 | | |
| Total (Corr.) | 2.83009E10 | 16 | | | |
| R-cuadrada = 99.4349 por ciento | | | | | |
| R-cuadrado (ajustado para g.l.) = 98.8698 por ciento | | | | | |
| Error estándar del est. = 4471.11 | | | | | |
| Error absoluto medio = 2603.18 | | | | | |
| Estadístico Durbin-Watson = 2.37667 (P=0.7526) | | | | | |
| Autocorrelación de residuos en retraso 1 = -0.193668 | | | | | |

2.2.3. Estudio de la Normalidad de los Residuos.

Para comprobar la normalidad de los residuos se dispone de dos caminos diferentes: por un lado está el gráfico de residuos, el cual ha de representar una campana de GAUSS perfecta o casi perfecta para que se pueda hablar de normalidad; y por otro lado el test para la normalidad de los residuos, en el cual se realizan varios test y se escoge la prueba con el más restrictivo, es decir, el test con el P.Valor más pequeño ya que es al que le costará más cumplir la hipótesis de normalidad H_0 .

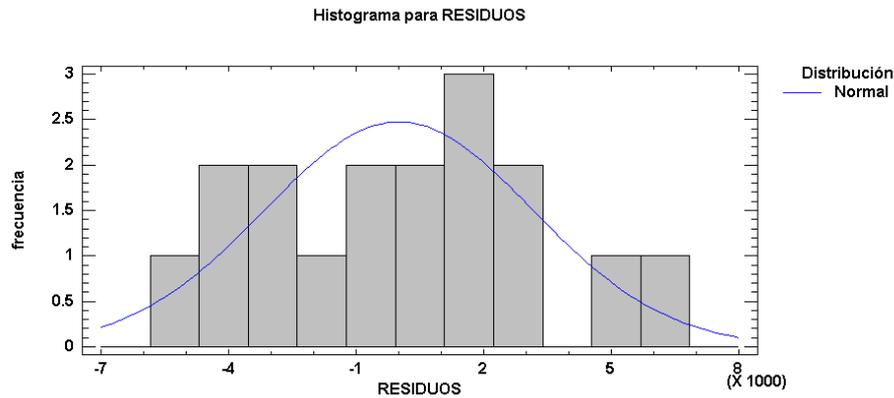


Figura IV. 18 Histograma de Residuos

Como se observa en el histograma, los residuos forman una campana de GAUSS casi perfecta, por lo que cabría esperar que los residuos se distribuyesen con normalidad (Figura IV. 18).

Tabla IV. 7. Pruebas de Normalidad.

| Prueba | Estadístico | Valor-P |
|-------------------------------|---------------------|----------|
| Chi-Cuadrado | 10.5294 | 0.309348 |
| Estadístico W de Shapiro-Wilk | 0.972655 | 0.837979 |
| Valor-Z para asimetría | 0.0668763 | 0.946675 |
| Valor-Z para curtosis | Datos Insuficientes | |

Se selecciona el P.Valor más pequeño de la Tabla IV. 7, en este caso el estadístico Chi-Cuadrado: P. Valor (0.309348) $>$ α (0,05). Como el P. valor es mayor que α , se Acepta H_0 . Esto concluye que los residuos se distribuyen con normalidad.

2.2.4. Heterocedasticidad

Para determinar la heterocedasticidad se calcula un modelo con los residuos al cuadrado y las variables explicativas del modelo que se está comprobando, obteniendo los siguientes resultados de la Tabla IV. 8, al realizar la estimación de los parámetros.

Tabla IV. 8 Resumen Regresión Múltiple para la comprobación de heterocedasticidad.

| | | <i>Error</i> | <i>Estadístico</i> | |
|------------------|-------------------|-----------------|--------------------|----------------|
| <i>Parámetro</i> | <i>Estimación</i> | <i>Estándar</i> | <i>T</i> | <i>Valor-P</i> |
| CONSTANTE | 5.34708E8 | 3.92379E8 | 1.36273 | 0.2101 |
| Becas_CCAA | -518.162 | 228.764 | -2.26504 | 0.0533 |
| Becas_Univ | 1249.91 | 228.149 | 5.47848 | 0.0006 |
| Becas_MICIN^4 | 2.59489E-10 | 3.83489E-10 | 0.676654 | 0.5177 |
| IPC | -414790. | 299524. | -1.38483 | 0.2035 |
| IPI | 29333.1 | 33480.4 | 0.876126 | 0.4065 |
| DESEMP_ESO | -846532. | 324275. | -2.61054 | 0.0311 |
| I+D+I | -14.6296 | 8.83485 | -1.65589 | 0.1363 |
| PARO | 274989. | 78171.7 | 3.51776 | 0.0079 |

En los P.Valores de los parámetros que acompañan a las variables explicativas se observa que hay tres variables que presentan problemas de heterocedasticidad (Becas_Univ, DESEMP-ESO, PARO). Se procederá con la solución del problema en otro apartado más adelante.

2.2.5. Autocorrelación

A continuación se representa el gráfico del FAS (Figura IV. 19) y el FAP (Figura IV. 20), para buscar la presencia de Autocorrelación. Como se puede observar en los gráficos, en ninguno de los dos, los retardos sobrepasan los límites, por lo que se puede afirmar que no hay problemas de Autocorrelación de ningún orden ni positiva ni negativa.

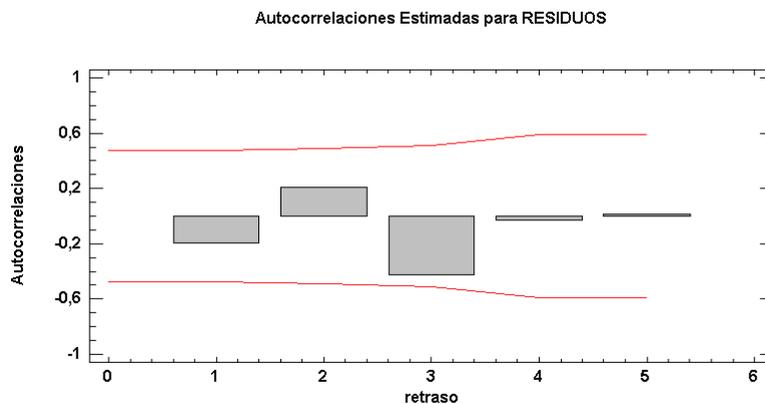


Figura IV. 19 Gráfico Función de autocorrelación simple.

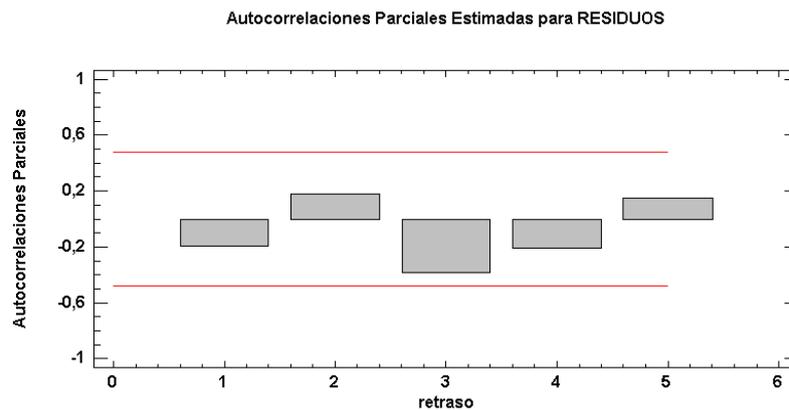


Figura IV. 20 Gráfico Función de autocorrelación parcial.

2.2.6. Residuos Atípicos y Puntos Influyentes.

En este punto se busca la presencia de puntos influyentes y residuos atípicos que puedan estar afectando a la formulación y comportamiento del modelo.

Tabla IV. 9 Puntos influyentes

| Fila | Influencia | Distancia de Mahalanobis | DFITS |
|------|------------|--------------------------|----------|
| 1 | 0.996566 | 4352.74 | 30.4301 |
| 4 | 0.615856 | 23.1103 | -1.86266 |
| 5 | 0.683569 | 31.4662 | -4.60071 |
| 9 | 0.902051 | 137.204 | -8.28127 |
| 10 | 0.58001 | 19.7776 | 3.28985 |
| 11 | 0.576209 | 19.4573 | -1.53663 |
| 13 | 0.973891 | 558.583 | 15.1247 |
| 16 | 0.759356 | 46.3952 | -3.19076 |

Influencia media de un solo punto = 0.529412

Ninguno de los valores de la columna de influencia (Tabla IV. 9) superan a la influencia media de un solo punto por 2 (1.058824), lo que significa que los puntos no son influyentes a priori. De los valores que se pueden ver en la columna DFITS (Tabla IV. 9), todos superan en valor absoluto el 1.45524074. Esto significa que los puntos son influyentes a posteriori. Los puntos influyentes a posterior representan a las siguientes CCAA: Andalucía, Baleares, Canarias, Cataluña, Comunidad Valenciana, Extremadura, Madrid y País Vasco.

Tabla IV. 10 Residuos Atípicos

| | | <i>Y</i> | | <i>Residuo</i> |
|-------------|----------|-----------------|----------------|----------------------|
| <i>Fila</i> | <i>Y</i> | <i>Predicha</i> | <i>Residuo</i> | <i>Estudentizado</i> |
| 5 | 16606.0 | 22039.0 | -5433.03 | -3.13 |
| 9 | 139014. | 141856. | -2841.56 | -2.73 |
| 10 | 71087.0 | 65130.6 | 5956.39 | 2.80 |
| 13 | 125868. | 124472. | 1396.37 | 2.48 |

Como se observa en la Tabla IV. 10 la fila 5 (que corresponde a la comunidad de Canarias) es superior a 3 en valor absoluto, por tanto, es un residuo atípico candidato a eliminar. Las filas 9, 10 y 13 (que corresponden a Cataluña, Comunidad Valenciana y Madrid respectivamente) son residuos atípicos ya que son mayores que 2 en valor absoluto. Con todas estas pruebas y habiendo detectado cuáles son los problemas a los que se enfrenta el modelo, se intentarán solucionar.

2.2.7. Solución problema Heterocedasticidad.

Se procederá a la eliminación de las variables que no son significativas, ya que hay 3 con heterocedasticidad, y alguna no es significativa, y es complicado solucionar el problema con más de una variable.

Tabla IV. 11 Resumen regresión múltiple

| | | <i>Error</i> | <i>Estadístico</i> | |
|------------------|-------------------|-----------------|--------------------|----------------|
| <i>Parámetro</i> | <i>Estimación</i> | <i>Estándar</i> | <i>T</i> | <i>Valor-P</i> |
| CONSTANTE | -126428. | 408216. | -0.309708 | 0.7647 |
| Becas_CCAA | -1.64701 | 0.237997 | -6.92027 | 0.0001 |
| Becas_Univ | 1.71096 | 0.237358 | 7.20835 | 0.0001 |
| Becas_MICIN^4 | 2.65209E-12 | 0 | 6.6474 | 0.0002 |
| IPC | 198.89 | 311.613 | 0.638261 | 0.5411 |
| IPI | -24.0771 | 34.8317 | -0.691242 | 0.5090 |
| DESEMP_ESO | -1256.96 | 337.363 | -3.72584 | 0.0058 |
| I+D+I | 0.108619 | 0.00919143 | 11.8174 | 0.0000 |
| PARO | 101.036 | 81.3267 | 1.24235 | 0.2493 |

En primer lugar, se elimina la variable PARO, ya que tiene el P.Valor más elevado (0,2493) de las que tienen problema de heterocedasticidad quedando el ajuste de parámetros como se observa en la Tabla IV. 12. A continuación se elimina la variable IPC, con un P.Valor de 0,8024 quedando el resumen de regresión múltiple como se observa en la Tabla IV. 13. Y por último eliminamos la variable IPI, con un P.Valor de 0,1745 quedando el ajuste como se observa en la Tabla IV. 14.

Se debería eliminar también la variable DESEMPLEO_ESO que tiene un P.Valor de 0,051 como se puede apreciar en la Tabla IV. 14, pero se va a dejar por el momento aunque no sea significativa por lo poco que supera el 0,05.

Tabla IV. 12 Resumen regresión múltiple sin la variable PARO

| | | <i>Error</i> | <i>Estadístico</i> | |
|------------------|-------------------|-----------------|--------------------|----------------|
| <i>Parámetro</i> | <i>Estimación</i> | <i>Estándar</i> | <i>T</i> | <i>Valor-P</i> |
| CONSTANTE | 230391. | 298713. | 0.771279 | 0.4603 |
| Becas_CCAA | -1.77187 | 0.222151 | -7.97597 | 0.0000 |
| Becas_Univ | 1.6772 | 0.242812 | 6.90741 | 0.0001 |
| Becas_MICIN^4 | 2.9567E-12 | 0 | 9.12305 | 0.0000 |
| IPC | -61.259 | 237.632 | -0.25779 | 0.8024 |
| IPI | -43.8359 | 31.9105 | -1.37371 | 0.2028 |
| DESEMP_ESO | -1271.63 | 347.186 | -3.66268 | 0.0052 |
| I+D+I | 0.102314 | 0.00789145 | 12.9652 | 0.0000 |

Tabla IV. 13 Resumen regresión múltiple sin la variable IPC

| | | <i>Error</i> | <i>Estadístico</i> | |
|------------------|-------------------|-----------------|--------------------|----------------|
| <i>Parámetro</i> | <i>Estimación</i> | <i>Estándar</i> | <i>T</i> | <i>Valor-P</i> |
| CONSTANTE | 154371. | 45344.2 | 3.40442 | 0.0067 |
| Becas_CCAA | -1.75471 | 0.201802 | -8.69517 | 0.0000 |
| Becas_Univ | 1.68727 | 0.228188 | 7.3942 | 0.0000 |
| Becas_MICIN^4 | 2.93843E-12 | 0 | 9.75824 | 0.0000 |
| IPI | -44.3348 | 30.3286 | -1.46181 | 0.1745 |
| DESEMP_ESO | -1269.72 | 330.508 | -3.84172 | 0.0033 |
| I+D+I | 0.102401 | 0.00750732 | 13.6401 | 0.0000 |

Tabla IV. 14 Resumen regresión múltiple sin la variable IPI

| | | <i>Error</i> | <i>Estadístico</i> | |
|------------------|-------------------|-----------------|--------------------|----------------|
| <i>Parámetro</i> | <i>Estimación</i> | <i>Estándar</i> | <i>T</i> | <i>Valor-P</i> |
| CONSTANTE | 100962. | 28209.6 | 3.579 | 0.0043 |
| Becas_CCAA | -1.59291 | 0.177246 | -8.987 | 0.0000 |
| Becas_Univ | 1.57201 | 0.224926 | 6.98899 | 0.0000 |
| Becas_MICIN^4 | 2.70029E-12 | 0 | 10.1508 | 0.0000 |
| DESEMP_ESO | -1196.06 | 343.109 | -3.48595 | 0.0051 |
| I+D+I | 0.108152 | 0.00671625 | 16.103 | 0.0000 |

Una vez eliminadas las variables de Paro, IPC y IPI se procede a guardar de nuevo los residuos y a realizar una nueva regresión múltiple con las variables significativas (14).

$$\text{RESIDUOS}^2 = \beta_0 + \beta_1 * \text{Beca CCAA} + \beta_2 * \text{Becas}_{\text{Univ}} + \beta_3 * \text{Beca MICIN}^4 + \beta_4 * \text{DESEMP ESO} + \beta_5 * (\text{I} + \text{D} + \text{I}) + \mu \quad (14)$$

Tabla IV. 15 Resumen de regresión múltiple con las variables significativas

| | | <i>Error</i> | <i>Estadístico</i> | |
|------------------|-------------------|-----------------|--------------------|----------------|
| <i>Parámetro</i> | <i>Estimación</i> | <i>Estándar</i> | <i>T</i> | <i>Valor-P</i> |
| CONSTANTE | 1.42151E8 | 8.44811E7 | 1.68263 | 0.1206 |
| Becas_CCAA | -562.096 | 530.809 | -1.05894 | 0.3123 |
| Becas_Univ | 2318.01 | 673.6 | 3.44123 | 0.0055 |
| Becas_MICIN^4 | 8.37319E-10 | 7.96658E-10 | 1.05104 | 0.3158 |
| DESEMP_ESO | -1.63611E6 | 1.02753E6 | -1.59227 | 0.1396 |
| I+D+I | -40.8631 | 20.1136 | -2.03162 | 0.0671 |

Se realiza de nuevo el test para identificar si existe problema de Heterocedasticidad con los residuos del modelo, y como se observa en la Tabla IV. 15 hay una variable que presenta problemas de Heterocedasticidad, ya que el P.Valor es menor que 0'05. Esta variable es la de Becas por Comunidades Autónomas para estudios universitarios. Como solo hay problemas en una variable (Becas_Univ), ahora sí que se procede a solucionar el problema de Heterocedasticidad.

Solución de Heterocedasticidad:

Se divide todo el modelo por la variable que da problemas de heterocedasticidad planteando el siguiente modelo (15):

$$\frac{\text{FLUJO MIG}}{\text{BECAS UNIV}^C} = \frac{\beta_0}{\text{BECAS UNIV}^C} + \frac{\text{BECAS CCAA}}{\text{BECAS UNIV}^C} + \frac{\text{BECAS UNIV}}{\text{BECAS UNIV}^C} + \frac{\text{BECAS MICIN}^4}{\text{BECAS UNIV}^C} + \frac{\text{DESEMP ESO}}{\text{BECAS UNIV}^C} + \frac{\text{I} + \text{D} + \text{I}}{\text{BECAS UNIV}^C} + \mu \quad (15)$$

Para el cálculo de C se procede siguiendo el método explicado en el apartado Heterocedasticidad (16). Como puede verse en la Tabla IV. 16, el valor de “h” encontrado es 0'8, por lo tanto C es 0'4, por lo que el modelo teórico para la solución de es el que se plantea en la ecuación 17.

$$\text{RESIDUOS}^2 = \beta_0 + \text{BECAS UNIV}^h + \mu \quad (16)$$

Tabla IV. 16. Búsqueda del valor de “h” para la solución de heterocedasticidad.

| h | R ² |
|------------|----------------|
| 1 | 30,4529 |
| 2 | 22,22 |
| -1 | 0,03481 |
| -2 | 0,03972 |
| -3 | 0,03971 |
| 0,5 | 28,10 |
| 0,4 | 25,93 |
| 0,6 | 29,529 |
| 0,7 | 30,36 |
| 0,8 | 30,75 |
| 0,9 | 30,73 |

$$\frac{\text{FLUJO MIG}}{\text{BECAS UNIV}^{0,4}} = \frac{1}{\text{BECAS UNIV}^{0,4}} + \frac{\text{BECAS CCAA}}{\text{BECAS UNIV}^{0,4}} + \frac{\text{BECAS UNIV}}{\text{BECAS UNIV}^{0,4}} + \frac{\text{BECAS MICIN}^4}{\text{BECAS UNIV}^{0,4}} + \frac{\text{DESEMP ESO}}{\text{BECAS UNIV}^{0,4}} + \frac{\text{I + D + I}}{\text{BECAS UNIV}^{0,4}} + \mu \quad (17)$$

Se elimina la constante que incorpora el programa por defecto, ya que esta se incluye en la regresión múltiple como $\frac{1}{\text{BECAS UNIV}^{0,4}}$, y el modelo estimado queda como se observa en la ecuación 18, y el resumen de la significación se observa en la Tabla IV. 17, dónde puede verse que todos los parámetros del modelo son significativos.

Tabla IV. 17 Modelo de regresión libre de Heterocedasticidad eliminando la constante

| | | <i>Error</i> | <i>Estadístico</i> | |
|--------------------------------|-------------------|-----------------|--------------------|----------------|
| <i>Parámetro</i> | <i>Estimación</i> | <i>Estándar</i> | <i>T</i> | <i>Valor-P</i> |
| 1/(Becas_Univ^0.4) | 63686.0 | 8699.29 | 7.32082 | 0.0000 |
| Becas_CCAA/(Becas_Univ^0.4) | -1.58241 | 0.217288 | -7.28256 | 0.0000 |
| Becas_Univ/(Becas_Univ^0.4) | 1.46227 | 0.445733 | 3.28059 | 0.0073 |
| Becas_MICIN^4/(Becas_Univ^0.4) | 2.70061E-12 | 0 | 6.98022 | 0.0000 |
| DESEMP_ESO/(Becas_Univ^0.4) | -737.087 | 118.25 | -6.23327 | 0.0001 |
| I+D+I/(Becas_Univ^0.4) | 0.109649 | 0.0127828 | 8.57788 | 0.0000 |

$$\frac{\text{Flujo_Migración}}{(\text{Becas_Univ}^{0.4})} = \frac{63686.0 * 1}{(\text{Becas_Univ}^{0.4})} - \frac{1.58241 * \text{Becas}_{\text{CCAA}}}{(\text{Becas_Univ}^{0.4})} +$$

$$\begin{aligned}
 & + \frac{1.46227 * \text{Becas}_{\text{Univ}}}{(\text{Becas}_{\text{Univ}}^{0.4})} + \frac{2.70061\text{E} - 12 * \text{Becas}_{\text{MICIN}}^4}{(\text{Becas}_{\text{Univ}}^{0.4})} - \frac{737.087 * \text{DESEMP_ESO}}{(\text{Becas}_{\text{Univ}}^{0.4})} \\
 & + \frac{0.109649 * I + D + I}{(\text{Becas}_{\text{Univ}}^{0.4})}
 \end{aligned}
 \tag{18}$$

Seguidamente se procede con la comprobación de la Heterocedasticidad del modelo solución estimado, y como se observa en la Tabla IV. 18 en esta modelo, ninguna de las variables presenta problemas de Heterocedasticidad ya que todos los P.Valores superan α (0.05).

Tabla IV. 18 Resumen regresión múltiple para la comprobación de la eliminación de Heterocedasticidad.

| <i>Parámetro</i> | <i>Estimación</i> | <i>Error Estándar</i> | <i>Estadístico T</i> | <i>Valor-P</i> |
|--------------------------------|-------------------|-----------------------|----------------------|----------------|
| CONSTANTE | -72569.2 | 189530. | -0.382891 | 0.7098 |
| 1/(Becas_Univ^0.4) | -1.09103E7 | 1.41823E7 | -0.76929 | 0.4595 |
| Becas_CCAA/(Becas_Univ^0.4) | -64.7972 | 102.367 | -0.632987 | 0.5409 |
| Becas_Univ/(Becas_Univ^0.4) | 192.574 | 528.629 | 0.364289 | 0.7232 |
| Becas_MICIN^4/(Becas_Univ^0.4) | 8.95842E-11 | 1.71908E-10 | 0.521118 | 0.6136 |
| DESEMP_ESO/(Becas_Univ^0.4) | 151768. | 196897. | 0.770796 | 0.4587 |
| I+D+I/(Becas_Univ^0.4) | -1.40394 | 5.27715 | -0.266041 | 0.7956 |

2.2.8. Comprobación

Una vez resueltos los problemas del modelo se comprueba el resto de características del modelo, empezando por la normalidad de los residuos.

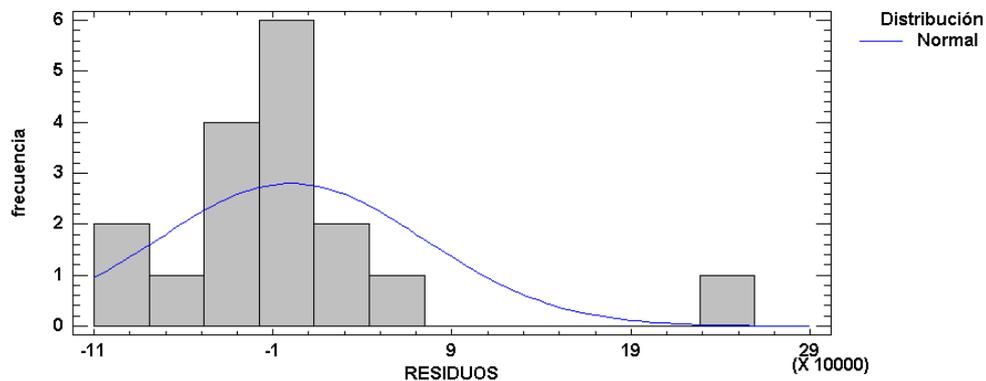


Figura IV. 21 Histograma de residuos.

Como se observa en la Figura IV. 21 no forma una campana de Gauss, por lo que cabría esperar que los residuos no siguiesen una distribución normal. Esto se corrobora al

observar que el P.Valor del test más restrictivo (Tabla IV. 19) es menor que 0'05. Con esto se afirma que los residuos no siguen una distribución normal. Cuando ocurre esto, se procede a ver si hay residuos atípicos que sean los causantes de la falta de distribución normal.

Tabla IV. 19 Pruebas de normalidad de los residuos

| <i>Prueba</i> | <i>Estadístico</i> | <i>Valor-P</i> |
|-------------------------------|---------------------|----------------|
| Chi-Cuadrado | 11.9412 | 0.216646 |
| Estadístico W de Shapiro-Wilk | 0.697921 | 0.0000527196 |
| Valor-Z para asimetría | 2.79603 | 0.00517355 |
| Valor-Z para curtosis | Datos Insuficientes | |

Como se puede ver en la Tabla IV. 20, la fila 14 (Murcia) es un residuo atípico candidato a eliminar, ya que es superior a 3. Se elimina la región de Murcia. Se vuelven a buscar la presencia de puntos anómalos, aparecen 2 pero ya no son candidatos a eliminar porque el residuo estudentizado está entre 2 i 3, como se observa en la Tabla IV. 21.

Tabla IV. 20 Residuos atípicos del modelo libre de Heterocedasticidad

| | <i>Y</i> | | <i>Residuo</i> | |
|-------------|----------|-----------------|----------------------|------|
| <i>Fila</i> | <i>Y</i> | <i>Predicha</i> | <i>Residuo</i> | |
| | | | <i>Estudentizado</i> | |
| 14 | 1712.35 | 1129.3 | 583.046 | 4.76 |

Tabla IV. 21 Residuos atípicos del modelo libre de Heterocedasticidad eliminando los datos de Murcia.

| | <i>Y</i> | | <i>Residuo</i> | |
|-------------|----------|-----------------|----------------------|------|
| <i>Fila</i> | <i>Y</i> | <i>Predicha</i> | <i>Residuo</i> | |
| | | | <i>Estudentizado</i> | |
| 8 | 582.061 | 348.278 | 233.784 | 2.08 |
| 10 | 1578.09 | 1376.89 | 201.201 | 2.16 |

Después de guardar los residuos del nuevo modelo, sin incluir la región de Murcia, se vuelven a realizar las pruebas de normalidad. En este caso como se contempla en la Figura IV. 22 el histograma forma una campana de Gauss casi perfecta, por lo que cabría esperar que los residuos se distribuyesen de forma normal. Seguidamente se realiza el test para la normalidad de los residuos. De los valores que aparecen en la Tabla IV. 22 se selecciona el P.Valor más pequeño, en este caso el estadístico W de Shapiro-Wilk: Puesto que el P. Valor (0.329536) > α (0,05) \rightarrow Aceptamos H_0 . Esto indica que los residuos siguen una distribución normal.

Tabla IV. 22 Pruebas para la normalidad de los residuos

| <i>Prueba</i> | <i>Estadístico</i> | <i>Valor-P</i> |
|-------------------------------|---------------------|----------------|
| Chi-Cuadrado | 9.5 | 0.392456 |
| Estadístico W de Shapiro-Wilk | 0.938653 | 0.329536 |
| Valor-Z para asimetría | 0.88412 | 0.37663 |
| Valor-Z para curtosis | Datos Insuficientes | |

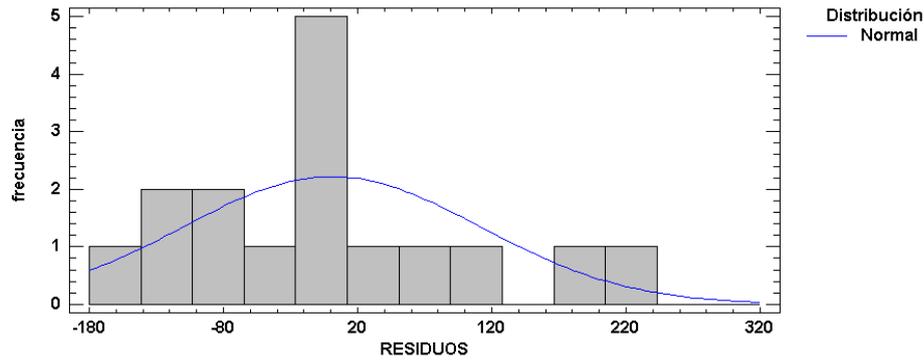


Figura IV. 22 Histograma de Residuos

Siguiendo con la comprobación de la validez del modelo propuesto, se realizan las pruebas de presencia de Autocorrelación. En los gráficos del FAS (Figura IV. 23) y FAP (Figura IV. 24), se observa que en ningún caso los factores de autocorrelación superan los límites de autocorrelación, por lo que el modelo no presenta problemas de esta clase.

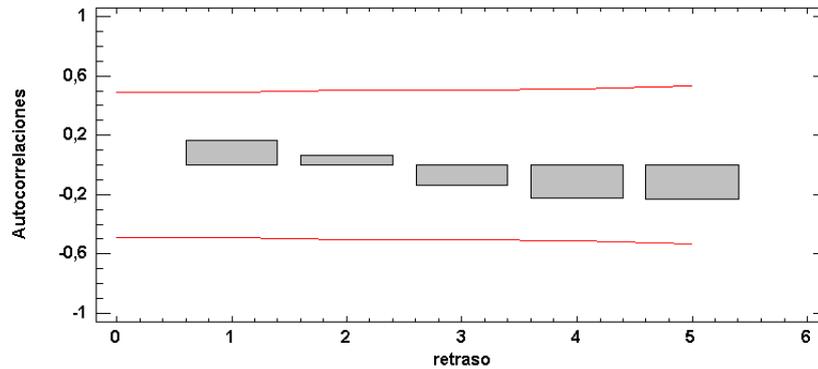


Figura IV. 23 Gráfico Función de autocorrelación simple del modelo libre de Heterocedasticidad y con distribución Normal de los residuos.

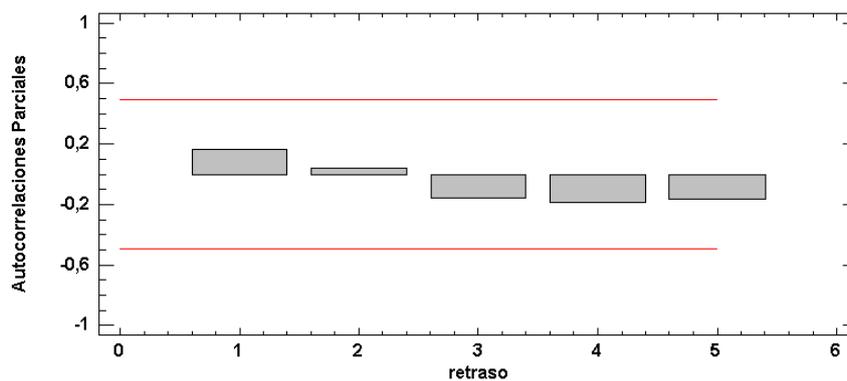


Figura IV. 24 Gráfico Función de autocorrelación parcial del modelo libre de Heterocedasticidad y con distribución Normal de los residuos.

3. Modelo Definitivo.

Después de realizar los cálculos y modificaciones pertinentes se obtiene el modelo de la ecuación (19). En este modelo, como se observa a continuación en la Tabla IV. 23, todos los parámetros son significativos. Como se aprecia en la Tabla IV. 24 el modelo en sí también es significativo y el R-cuadrado es de 99'9349, lo que significa que estas variables explican el modelo en un 99.93 %. Esto es un porcentaje muy elevado e indica que el modelo es muy válido.

$$\begin{aligned} \frac{\text{Flujo}_{\text{Migración}}}{(\text{Becas}_{\text{Univ}}^{0.4})} = & \frac{74301.5 * 1}{(\text{Becas}_{\text{Univ}}^{0.4})} - \frac{1.52553 * \text{Becas}_{\text{CCAA}}}{(\text{Becas}_{\text{Univ}}^{0.4})} + \\ & + \frac{1.66515 * \text{Becas}_{\text{Univ}}}{(\text{Becas}_{\text{Univ}}^{0.4})} + \frac{2.64248E - 12 * \text{Becas}_{\text{MICIN}}^4}{(\text{Becas}_{\text{Univ}}^{0.4})} - \\ & - \frac{883.656 * \text{DESEMP_ESO}}{(\text{Becas}_{\text{Univ}}^{0.4})} + \frac{0.110383 * (I + D + I)}{(\text{Becas}_{\text{Univ}}^{0.4})} \end{aligned}$$

(19)

Tabla IV. 23 Resumen regresión múltiple del modelo definitivo

| Parámetro | Estimación | Error | | Estadístico | Valor-P |
|--------------------------------|-------------|------------|----------|-------------|---------|
| | | Estándar | T | | |
| 1/(Becas_Univ^0.4) | 74301.5 | 5516.62 | 13.4687 | 0.0000 | |
| Becas_CCAA/(Becas_Univ^0.4) | -1.52553 | 0.126612 | -12.0489 | 0.0000 | |
| Becas_Univ/(Becas_Univ^0.4) | 1.66515 | 0.262052 | 6.35428 | 0.0001 | |
| Becas_MICIN^4/(Becas_Univ^0.4) | 2.64248E-12 | 0 | 11.7566 | 0.0000 | |
| DESEMP_ESO/(Becas_Univ^0.4) | -883.656 | 75.182 | -11.7536 | 0.0000 | |
| I+D+I/(Becas_Univ^0.4) | 0.110383 | 0.00741682 | 14.8829 | 0.0000 | |

Tabla IV. 24 Resumen análisis regresión múltiple

| Fuente | Suma de Cuadrados | Gl | Cuadrado Medio | Razón-F | Valor-P |
|--|-------------------|----|----------------|---------|---------|
| Modelo | 2.81468E8 | 6 | 4.69114E7 | 2556.90 | 0.0000 |
| Residuo | 183469. | 10 | 18346.9 | | |
| Total | 2.81652E8 | 16 | | | |
| R-cuadrada = 99.9349 por ciento | | | | | |
| R-cuadrado (ajustado para g.l.) = 99.9023 por ciento | | | | | |
| Error estándar del est. = 135.451 | | | | | |
| Error absoluto medio = 79.7849 | | | | | |
| Estadístico Durbin-Watson = 1.60921 | | | | | |
| Autocorrelación de residuos en retraso 1 = 0.166244 | | | | | |

3.1. Interpretación de los parámetros y predicción.

En este modelo a la hora de interpretar los parámetros podemos observar que resulta muy complicado. Por esto se procede a hacer una predicción modificando cada una de las variables drásticamente para poder interpretar el modelo. Se modifica cada una de las variables bajando drásticamente a 1, para ver cómo reacciona el modelo y poder sacar conclusiones (Figura IV. 25).

| Predicción | Becas_CCAA | Becas_Univ | Becas_MICIN | IPC | IPI | DESEMP_ESO | I+D+I | PARO |
|------------|------------|------------|-------------|----------|----------|------------|-------|-------|
| 1 | 1 | 385,1 | 464,5 | 1246,318 | 1045,007 | 80,22 | 14993 | 80,14 |
| 2 | 101,5 | 1 | 464,5 | 1246,318 | 1045,007 | 80,22 | 14993 | 80,14 |
| 3 | 101,5 | 385,1 | 1 | 1246,318 | 1045,007 | 80,22 | 14993 | 80,14 |
| 4 | 101,5 | 385,1 | 464,5 | 1 | 1045,007 | 80,22 | 14993 | 80,14 |
| 5 | 101,5 | 385,1 | 464,5 | 1246,318 | 1 | 80,22 | 14993 | 80,14 |
| 6 | 101,5 | 385,1 | 464,5 | 1246,318 | 1045,007 | 1 | 14993 | 80,14 |
| 7 | 101,5 | 385,1 | 464,5 | 1246,318 | 1045,007 | 80,22 | 1 | 80,14 |
| 8 | 101,5 | 385,1 | 464,5 | 1246,318 | 1045,007 | 80,22 | 14993 | 1 |

Figura IV. 25 Tabla de modificación de los valores de las X para la obtención de predicciones.

| Predicción | Variable mod | Ajustado | Error Est. | Inferior 95.0% | Superior 95.0% | Inferior 95.0% |
|------------|--------------|----------|--------------------|--------------------|--------------------|------------------|
| | | | LC para Pronóstico | LC para Pronóstico | LC para Pronóstico | LC para la Media |
| 1 | Becas_CCAA | 5709 | 1558 | 2238 | 9181 | 4531 |
| 2 | Becas_Univ | 53197 | 6220 | 39337 | 67057 | 39728 |
| 3 | Becas_MICIN | 5556 | 1557 | 2086 | 9026 | 4383 |
| 4 | IPC | 5556 | 1557 | 2086 | 9026 | 4383 |
| 5 | IPI | 5556 | 1557 | 2086 | 9026 | 4383 |
| 6 | DESEMP_ESO | 75559 | 5669 | 62928 | 88191 | 63357 |
| 7 | I+D+I | 3901 | 1568 | 408 | 7395 | 2660 |
| 8 | PARO | 5556 | 1557 | 2086 | 9026 | 4383 |

Figura IV. 26 Resultado de las predicciones en función de las modificaciones de las X, para $\text{Flujo_Migración}/(\text{Becas_Univ}^{0.4})$.

Debido a que el modelo propuesto inicialmente tenía algunos problemas se han tenido que modificar algunas variables para que estos problemas desapareciesen. Uno de los cambios a sido elevar a 4 la variable "Becas_Univ" y dividir todas las variables del modelo por esta misma. Llegados a este punto, eso se deberá tener en cuenta a la hora de realizar los cálculos. Para obtener el valor de la variable dependiente despejado, se ha tenido que multiplicar el resultado de cada predicción por la variable "Becas_Univ⁴" (20), con lo que se obtienen los resultados que se muestran en la Figura IV. 26.

$$\text{Flujo migratorio} = \text{predicción} * \text{Becas_Univ}^{0.4} \quad (20)$$

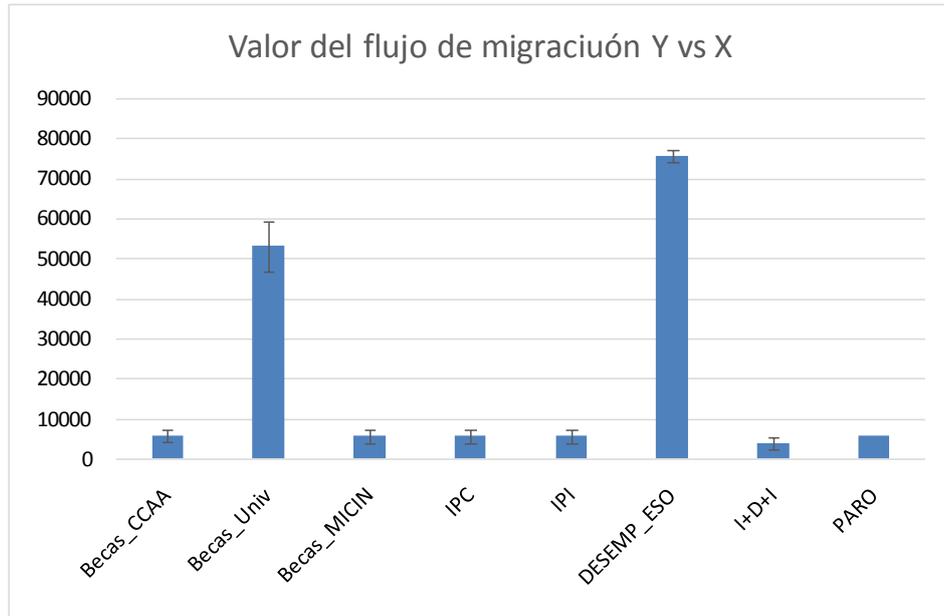


Figura IV. 27 Representación del valor de las predicciones para el flujo de migración en función de la variable X modificada.

Tras analizar los resultados de las predicciones y como se observa en la Figura IV. 27 existen dos variables que destacan sobre las demás. Estas variables son Becas_Univ y Desempleo_ESO. Tras realizar este modelo se puede llegar a la conclusión de que ambas variables son las que más influencia ejercen sobre el flujo de migración. Por lo tanto, si disminuye el número de becas universitarias o el de desempleo eso aumenta el flujo de migración. Esto puede explicarse en el caso de las becas debido a que los estudiantes que no reciben becas deciden buscar formación o empleo en el extranjero.

El caso más difícil de explicar es el del Desempleo_ESO puesto que cabría esperar que a mayor desempleo de jóvenes se diera mayor flujo de emigración, y sin embargo la influencia es al contrario. Probablemente porque estos jóvenes optan más por continuar su formación en su país de origen antes que emigrar a otro país.

4. Planteamiento Segundo Modelo.

4.1. Modelo teórico.

Como se ha podido observar en el modelo anterior, la interpretación de los parámetros cuando se incluyen tantas variables puede ser muy complicada, sobre todo cuando se realizan alteraciones en la estructura lineal. Por este motivo se decide buscar un modelo lineal más sencillo reduciendo el número de variables explicativas. En este punto se observó que la variable I+D+I explicaba por sí sola más del 80% del comportamiento del flujo de emigración. Entonces se decidió plantear el siguiente modelo teórico de la ecuación 21. Todo esto, con la misma matriz de datos utilizada en el primer modelo propuesto (Figura IV. 5).

$$Flujo_{Migración} = \beta_0 + \beta_1 * (I + D + I) + U \quad (21)$$

4.2. Estimación de la ecuación de regresión múltiple y validación del modelo.

En la ecuación 22 se representa el modelo teórico anterior con los parámetros estimados. En los siguientes apartados del proyecto se procederá a validar el modelo paso a paso como se indica en el apartado 3.2 de Metodología.

$$Flujo_{Migración} = 7363.99 + 0.165701 * (I + D + I) \quad (22)$$

4.2.1. Análisis de la significación

En este apartado se procederá a identificar si la variable explicativa es significativa o no, así como el modelo en sí. Como se observa en la Tabla IV. 25 el P.Valor de la variable I+D+I es menor que 0'05, lo que indica que la variable es significativa. En cambio la constante es superior a 0'05, por tanto, no es significativa para el modelo, con lo que se deberá eliminar como se verá más adelante. En la Tabla IV. 26 se contempla que el P.Valor del modelo es 0'000, lo que muestra que el modelo en sí es significativo. Además se puede ver a través de el R² que solo esta variable explica un 85'94% del modelo, un dato muy elevado para tratarse de únicamente una variable.

Tabla IV. 25 Resumen análisis regresión múltiple

| Parámetro | Estimación | Error | | Estadístico | Valor-P |
|-----------|------------|-----------|---------|-------------|---------|
| | | Estándar | T | | |
| CONSTANTE | 7363.99 | 4663.66 | 1.57901 | | 0.1352 |
| IDI | 0.165701 | 0.0173026 | 9.57662 | | 0.0000 |

Tabla IV. 26 Resumen regresión múltiple

| Fuente | Suma de Cuadrados | Gl | Cuadrado Medio | Razón-F | Valor-P |
|--|-------------------|----|----------------|---------|---------|
| Modelo | 2.43228E10 | 1 | 2.43228E10 | 91.71 | 0.0000 |
| Residuo | 3.97814E9 | 15 | 2.65209E8 | | |
| Total (Corr.) | 2.83009E10 | 16 | | | |
| R-cuadrada = 85.9434 por ciento | | | | | |
| R-cuadrado (ajustado para g.l.) = 85.0063 por ciento | | | | | |
| Error estándar del est. = 16285.3 | | | | | |
| Error absoluto medio = 9969.73 | | | | | |
| Estadístico Durbin-Watson = 1.21396 (P=0.0462) | | | | | |
| Autocorrelación de residuos en retraso 1 = 0.382689 | | | | | |

Para solucionar el problema que provoca la constante al no ser significativa se utiliza un método en el cual tanto a la variable I+D+I como a la constante se le resta la media de ellas mismas, es decir, en el caso de la variable explicada se procede con el cálculo de $Y - \bar{Y}$, y en el de la variable explicativa con $X - \bar{X}$. En la Figura IV. 28 se puede ver gráficamente que es lo que ocurre cuando se realiza esta modificación.

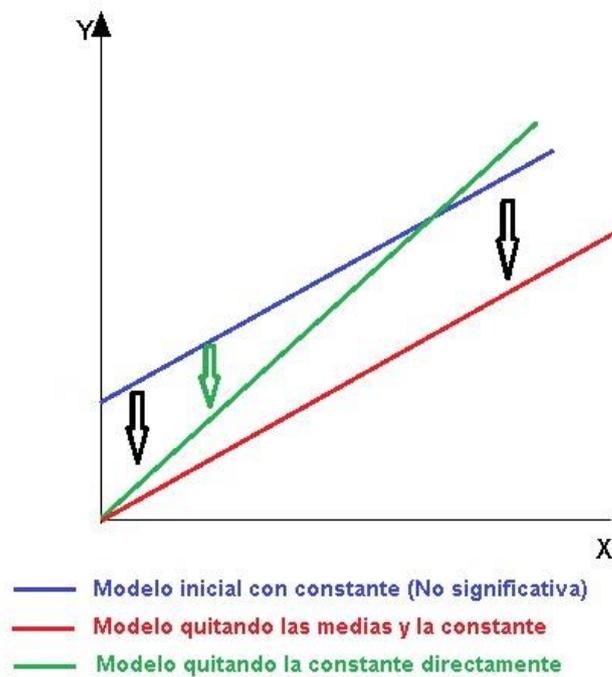


Figura IV. 28 Gráfico explicativo eliminación de la constante

Tras realizar los cálculos para intentar solucionar el problema, el programa modifica el modelo teórico estimado, quedando este como se observa en la ecuación 23. Con este nuevo modelo se proceden a realizar todas las pruebas de validación.

En primer lugar se vuelve a comprobar la significación. Tanto la variable I+D+I como el modelo en sí son significativos, como se observa en la Tabla IV. 27 y la Tabla IV. 28. Además, en la Tabla IV. 28 se puede ver que el R² sigue siendo de 85'9434.

$$\text{Flujo Emigración} - \text{AVG}(\text{Flujo Emigración}) = 0.165701 * \text{IDI} - \text{AVG}(\text{IDI}) \quad (23)$$

Tabla IV. 27 Resumen regresión múltiple de la variable I+D+I

| | | Error | Estadístico | |
|--------------|------------|-----------|-------------|---------|
| Parámetro | Estimación | Estándar | T | Valor-P |
| IDI-AVG(IDI) | 0.165701 | 0.0167532 | 9.89069 | 0.0000 |

Tabla IV. 28 Resumen regresión múltiple segundo modelo

| Fuente | Suma de Cuadrados | Gl | Cuadrado Medio | Razón-F | Valor-P |
|--|-------------------|----|----------------|---------|---------|
| Modelo | 2.43228E10 | 1 | 2.43228E10 | 97.83 | 0.0000 |
| Residuo | 3.97814E9 | 16 | 2.48634E8 | | |
| Total | 2.83009E10 | 17 | | | |
| R-cuadrada = 85.9434 por ciento | | | | | |
| R-cuadrado (ajustado para g.l.) = 85.9434 por ciento | | | | | |
| Error estándar del est. = 15768.1 | | | | | |
| Error absoluto medio = 9969.73 | | | | | |
| Estadístico Durbin-Watson = 1.21396 | | | | | |
| Autocorrelación de residuos en retraso 1 = 0.382689 | | | | | |

4.2.2. Análisis de los gráficos de residuos

En este apartado se muestran los gráficos donde se comparan los residuos frente a cada una de variables explicativas con el fin de encontrar si existe algún tipo de problema que seguidamente se tenga que corregir.

Como se puede ver en la Figura IV. 29 y Figura IV. 30, no se observan formas en los residuos que se puedan asociar a problemas. Lo único que se pueden apreciar son algunos residuos atípicos que pueden ser candidatos a eliminar. Ocurre lo mismo en la Figura IV. 31, donde los residuos se ordenan aleatoriamente por lo que no cabría esperar que pueda haber autocorrelación de primer orden positiva ni negativa.

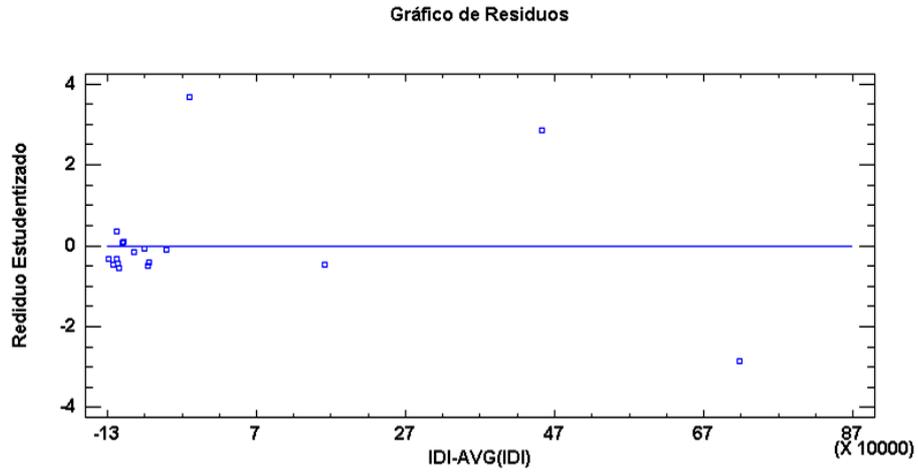


Figura IV. 29 Gráfico de residuos vs variable I+D+I

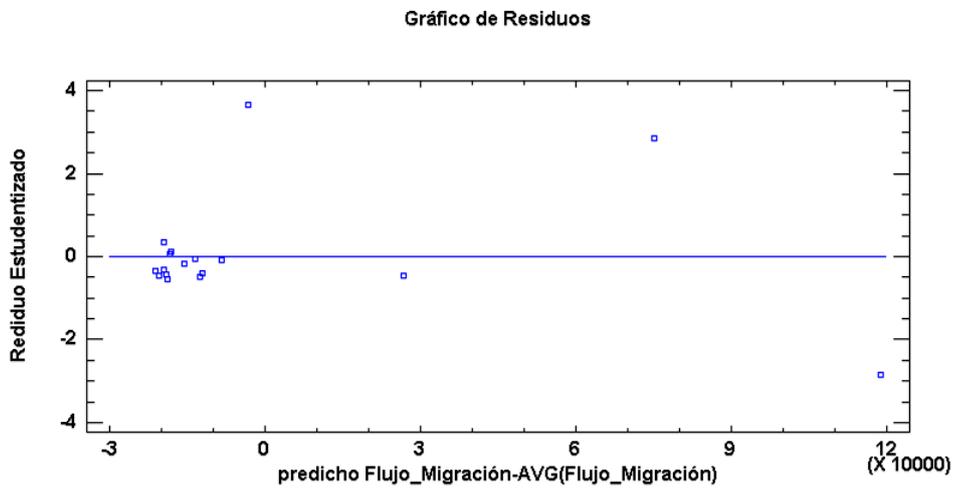


Figura IV. 30 Gráfico de residuos frente al flujo de emigración.

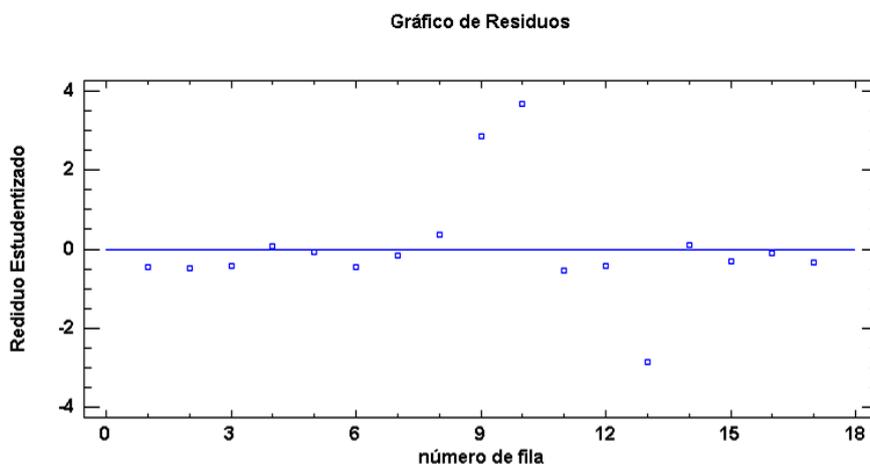


Figura IV. 31 Gráfico de residuos frente a número de fila.

4.2.3. Estudio de la normalidad de los residuos

En este apartado se procederá a realizar las pruebas gráficas y el test para la normalidad de los residuos. Como se observa en la Figura IV. 32 los residuos no forman una campana de gauss perfecta, por lo que cabría esperar que los residuos no se distribuyesen con normalidad. A continuación se selecciona el P.Valor más restrictivo de la Tabla IV. 29, en este caso, el Estadístico W de Shapiro-Wilk (0'000318517). Como este valor es inferior a 0'05 se concluye que los residuos no se distribuyen con normalidad. Esto podría deberse a la existencia de puntos atípicos.

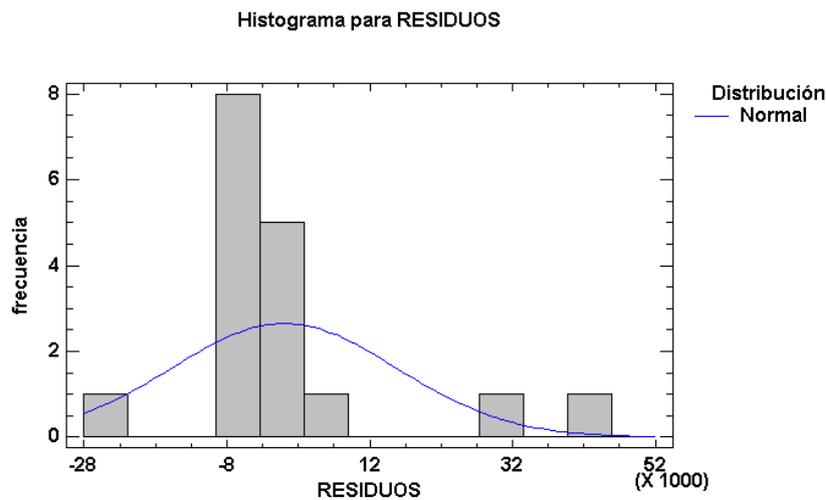


Figura IV. 32 Histograma de residuos del segundo modelo planteado.

Tabla IV. 29 Test para la normalidad de los residuos.

| Prueba | Estadístico | Valor-P |
|-------------------------------|---------------------|-------------|
| Chi-Cuadrado | 20.4118 | 0.0155345 |
| Estadístico W de Shapiro-Wilk | 0.753631 | 0.000318517 |
| Valor-Z para asimetría | 2.09195 | 0.0364428 |
| Valor-Z para curtosis | Datos Insuficientes | |

4.2.4. Residuos atípicos y puntos influyentes

En el análisis de los gráficos de los residuos, se observaba gráficamente la existencia de algunos residuos atípicos. A continuación se realizan las pruebas pertinentes para comprobar si lo que se observa de forma gráfica coincide con la realidad. En la Tabla IV. 30 se observa que la fila 10 (que corresponde a la comunidad Valenciana) es superior a 3 en valor absoluto, por tanto, es un residuo atípico candidato a eliminar. Las filas 9 y 13

(que corresponden a Cataluña y Madrid respectivamente) son residuos atípicos ya que son mayores que 2 en valor absoluto. Numéricamente solo sería necesario eliminar la variable de la fila 10, pero en este caso se decide eliminar las tres variables debido a su alto residuo estudentizado, y a que se valora que estas tres comunidades pueden estar distorsionando el resto de los resultados. Tras eliminar estas tres filas, se procederá a realizar de nuevo todas las pruebas de validación del nuevo modelo sin estas tres comunidades.

Tabla IV. 30 Residuos atípicos segundo modelo propuesto

| | | <i>Y</i> | | <i>Residuo</i> |
|-------------|----------|-----------------|----------------|----------------------|
| <i>Fila</i> | <i>Y</i> | <i>Predicha</i> | <i>Residuo</i> | <i>Estudentizado</i> |
| 9 | 107902. | 75107.8 | 32794.5 | 2.85 |
| 10 | 39975.3 | -3371.18 | 43346.5 | 3.67 |
| 13 | 94756.3 | 118899. | -24142.9 | -2.84 |

4.3. Modelo teórico sin las comunidades de Cataluña, Comunidad Valenciana y Madrid.

En la ecuación 24 se representa el modelo teórico con los parámetros estimados. En los siguientes apartados del proyecto se procederá a validar el modelo paso a paso como se indica en el apartado 3.2 de Metodología.

$$\text{Flujo Emigración} - \text{AVG}(\text{Flujo Emigración}) = 0.151987 * \text{IDI} - \text{AVG}(\text{IDI}) \quad (24)$$

4.3.1. Análisis de la significación

Al realizar el ajuste del modelo sin las tres comunidades autónomas que generaban los residuos atípicos, se observa que sigue siendo significativa la variable y el propio modelo, y que el ajuste ha aumentado hasta 87,17% como se puede ver en la Tabla IV. 31 y la Tabla IV. 32.

Tabla IV. 31 Resumen regresión múltiple

| | | <i>Error</i> | <i>Estadístico</i> | |
|------------------|-------------------|-----------------|--------------------|----------------|
| <i>Parámetro</i> | <i>Estimación</i> | <i>Estándar</i> | <i>T</i> | <i>Valor-P</i> |
| IDI-AVG(IDI) | 0.151987 | 0.0161759 | 9.39589 | 0.0000 |

Tabla IV. 32 Resumen regresión múltiple

| Fuente | Suma de Cuadrados | Gl | Cuadrado Medio | Razón-F | Valor-P |
|--|-------------------|----|----------------|---------|---------|
| Modelo | 1.63535E9 | 1 | 1.63535E9 | 88.28 | 0.0000 |
| Residuo | 2.40812E8 | 13 | 1.8524E7 | | |
| Total | 1.87616E9 | 14 | | | |
| R-cuadrada = 87.1646 por ciento | | | | | |
| R-cuadrado (ajustado para g.l.) = 87.1646 por ciento | | | | | |
| Error estándar del est. = 4303.95 | | | | | |
| Error absoluto medio = 3606.42 | | | | | |
| Estadístico Durbin-Watson = 2.93971 | | | | | |
| Autocorrelación de residuos en retraso 1 = -0.471444 | | | | | |

4.3.2. Estudio de la normalidad de los residuos

Una vez eliminadas las comunidades autónomas pertinentes y comprobada la significación del modelo se procede con la comprobación de la normalidad de los residuos. Como se observa en la Figura IV. 33 los residuos forman una campana de Gauss casi perfecta, por lo que cabría esperar que siguiesen una distribución normal. Esto se corrobora al observar que el P.Valor del test más restrictivo (Tabla IV. 33) es mayor que 0'05. Con esto se afirma que los residuos siguen una distribución normal.

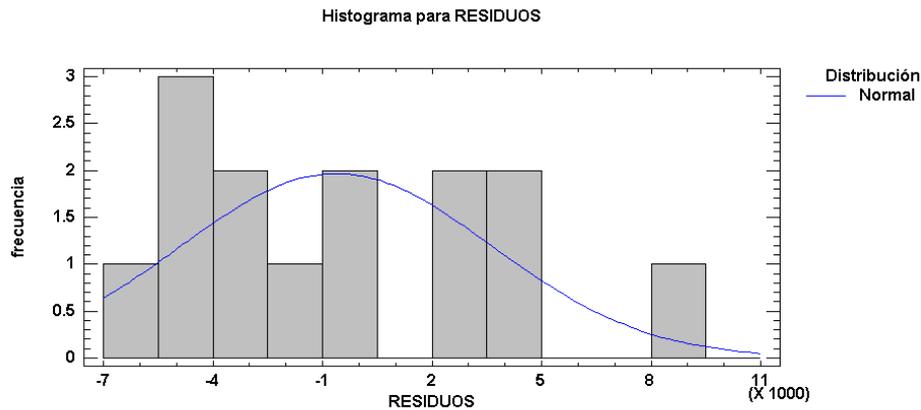


Figura IV. 33 Histograma de residuos

Tabla IV. 33 Pruebas de normalidad de los residuos

| Prueba | Estadístico | Valor-P |
|-------------------------------|---------------------|----------|
| Chi-Cuadrado | 8.0 | 0.43347 |
| Estadístico W de Shapiro-Wilk | 0.942056 | 0.425702 |
| Valor-Z para asimetría | 0.815635 | 0.414707 |
| Valor-Z para curtosis | Datos Insuficientes | |

4.3.3. Heterocedasticidad

Para determinar si existen problemas de heterocedasticidad, se realiza un modelo de residuos al cuadrado, como se ha realizado anteriormente. La Tabla IV. 34 muestra el resultado de la estimación de los parámetros del modelo de los residuos al cuadrado frente a la variable explicativa. Como el P.Valor de esta variable I+D+I es superior a 0'05 se puede confirmar que este modelo está libre de Heterocedasticidad.

Tabla IV. 34 Resumen Regresión Múltiple para la comprobación de heterocedasticidad.

| | | <i>Error</i> | <i>Estadístico</i> | |
|------------------|-------------------|-----------------|--------------------|----------------|
| <i>Parámetro</i> | <i>Estimación</i> | <i>Estándar</i> | <i>T</i> | <i>Valor-P</i> |
| CONSTANTE | 2.26621E7 | 6.28302E6 | 3.60688 | 0.0036 |
| IDI | -89.3235 | 67.0605 | -1.33198 | 0.2076 |

4.3.4. Autocorrelación

Para comprobar la existencia de autocorrelación en este trabajo se emplean los gráficos del FAS (Figura IV. 34) y el FAP (Figura IV. 35). En ambos casos se observa que el factor de autocorrelación de ningún retardo supera los límites de autocorrelación, con lo que se comprueba que no existe autocorrelación de ningún orden positiva ni negativa.

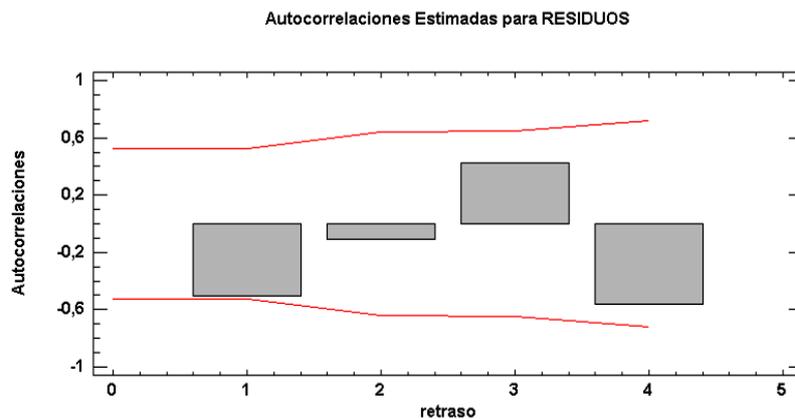


Figura IV. 34 Función de autocorrelación simple (FAS)

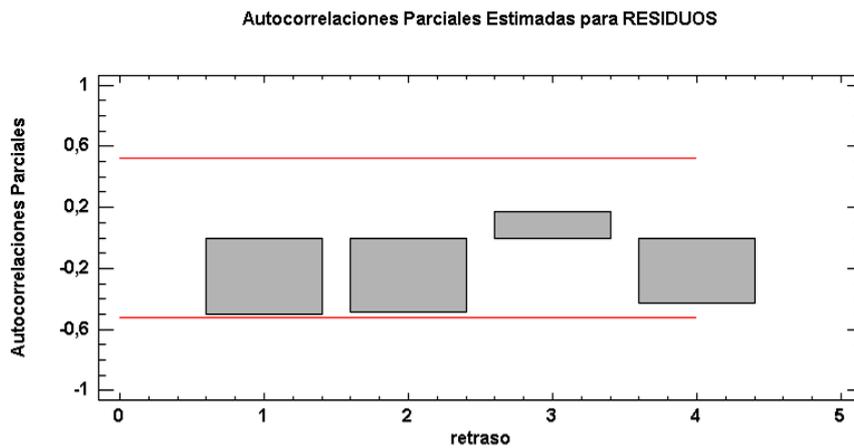


Figura IV. 35 Función de autocorrelación Parcial (FAP)

4.3.5. Residuos atípicos y puntos influyentes

En este punto se analizan los residuos atípicos que puede tener el modelo. Se puede ver en la Tabla IV. 35 que solo existe un residuo atípico que corresponde a la comunidad de Castilla la mancha, pero como no supera a 3 en valor absoluto no se procederá a su eliminación.

Tabla IV. 35 Residuos atípicos

| | | <i>Y</i> | <i>Residuo</i> | |
|-------------|----------|-----------------|----------------|----------------------|
| <i>Fila</i> | <i>Y</i> | <i>Predicha</i> | <i>Residuo</i> | <i>Estudentizado</i> |
| 6 | 3513.29 | -4810.05 | 8323.34 | 2.22 |

4.3.6. Interpretación de los parámetros y Predicción

En este punto se procede a la interpretación de los parámetros del modelo así como a la predicción. En primer lugar, en el segundo modelo propuesto se ha eliminado la constante, esto deberá tenerse en cuenta a la hora de interpretarlo. Dado que se le ha restado la media, tanto a la Y como a la variable X, al realizar la predicción deberá de sumarse el promedio de todas las predicciones que se calculen ($\bar{Y} = 13.780'7143$). Para realizar esta predicción se descende el valor de la variable I+D+I drásticamente a 1, con lo que se obtienen los datos que aparecen en la Tabla IV. 36. Al tratarse de movimientos migratorios los resultados no pueden tener decimales al igual que no pueden ser negativos, por este motivo, se ha redondeado las cifras decimales en la predicción y se han sustituido los números negativos por 0.

Tabla IV. 36 Resultado de la predicción

| | Ajustado | Error Est. | Inferior 95.0% | Superior 95.0% |
|---------------|----------|--------------------|--------------------|--------------------|
| Fila | | LC para Pronóstico | LC para Pronóstico | LC para Pronóstico |
| Programa (15) | -8672,81 | 4401,82 | -18182,40 | 836,76 |
| Predicción Y | 5108 | 18183 | 0 | 14617 |

Los resultados que se observan en la Tabla IV. 36 nos hacen pensar que por mucho que disminuya la variable I+D+I no se evita la salida masiva de emigrantes españoles. Cuanta más inversión en I+D+I, mayor será la emigración, contrariamente a lo que se esperaba. Esto podría estar ocurriendo porque parte de estas ayudas se destinen a formar a los trabajadores de las empresas en el extranjero, ya que en parte con estas ayudas las empresas tratan de expandirse y abrir nuevos mercados, aunque este tema necesitaría un análisis más profundo que desvelara los motivos de este fenómeno.

Si se compara la predicción de la variable I+D+I del primer modelo con el segundo (Tabla IV. 37), se puede observar que en el segundo modelo el valor predicho es más elevado, pero el intervalo de confianza también es más amplio, por lo que el valor de la predicción será más impreciso en este segundo caso. Esto se debe a la diferencia de ajuste entre ambos modelos, puesto que en el primero superaba el 99%, y en este segundo modelo no se alcanza el 90%.

Tabla IV. 37 Comparación de la predicción del primer modelo con el segundo modelo.

| I+D+I | Ajustado | Error Est. | Inferior 95.0% | Superior 95.0% |
|----------------------|----------|--------------------|--------------------|--------------------|
| Fila | | LC para Pronóstico | LC para Pronóstico | LC para Pronóstico |
| Programa (15) | -8672,81 | 4401,82 | -18182,40 | 836,76 |
| Predicción 2º Modelo | 5108 | 18183 | 0 | 14617 |
| Predicción 1º Modelo | 3901 | 1568 | 408 | 7395 |

V. Conclusiones

En el siguiente apartado se procede con el listado de las conclusiones obtenidas en este trabajo:

- Se ha conseguido realizar la modelización del flujo de emigración de los españoles por comunidades autónomas en los últimos tiempos.
- Se han conseguido determinar cuáles son los factores más relevantes en la emigración de los españoles. Los factores realmente significativos han sido: Becas_Univ, Becas_CCAA, Becas MICIN, DESEMPLEO_ESO, I+D+I (este factor es capaz de explicar el 80% del valor de los emigrantes).
- Se han planteado y validado dos modelos de regresión múltiple lineal para explicar el fenómeno de la emigración en España. Con ambos se han realizado predicciones con valores hipotéticos de las variables explicativas.
- De los dos modelos planteados se recomendaría el uso del primer modelo, en el que se incluyen las variables: Becas_Univ, Becas_CCAA, Becas_MICIN, DESEMPLEO_ESO, I+D+I. La razón principal es que en este modelo se incluyen los datos de mas comunidades autónomas y el intervalo de la predicción es menor, es decir, el error que se comete en la predicción es más pequeño que si se emplea el modelo planteado solo con la variable I+D+I. Sin embargo, se debe recalcar que este último modelo es igualmente valido y podría ser una herramienta en caso de no disponer de todas las variables X incluidas en el modelo anterior.

VI. Futuras líneas de investigación

En este apartado se procede a listar las posibles futuras líneas de investigación que surgen de este proyecto:

- Una posible alternativa sería el planteamiento de modelos de regresión múltiple no lineal con las mismas variables analizadas en este trabajo.
- Se podría modelizar el comportamiento por separado de las comunidades de Cataluña, Comunidad Valenciana y Madrid, ya que en ambos modelos aparecen como residuos atípicos y puntos influyentes lo cual indica que estas comunidades tienen un comportamiento diferente al resto.
- Buscar otro tipo de variables de entorno socio-económicas que puedan explicar el flujo de emigración de los españoles.
- Otra alternativa sería realizar encuestas para determinar las causas de emigración directa de los españoles. Recogiendo opiniones, vivencias, etc.

VII. Bibliografía

[1]<https://es.wikipedia.org/wiki/Emigraci%C3%B3n>

[2]<http://www.definicionabc.com/general/emigrante.php>

[3]<http://quees.la/emigracion/>

[4]<http://retornoasistido.blogspot.com.es/2010/11/breve-historia-de-la-emigracion.html>

[5]Ministerio de Trabajo y asuntos sociales (2008) <<Análisis de las Migraciones>>, ISBN: 978-84-691-1789-7, paginas (37-42), Ed. ACCEM, España.

[6]Periódico online el País. URL:

http://economia.elpais.com/economia/2013/11/07/empleo/1383838356_819951.html

[7]Periódico online el País. URL:

http://ccaa.elpais.com/ccaa/2013/10/08/valencia/1381232397_099666.html

[8]<http://www.nonosvamosnosechan.net/>

[9]<http://erafbadia.blogspot.com.es/>

[10]<http://www.rtve.es/television/espanoles-en-el-mundo/>

[11]<https://es-es.facebook.com/>

[12]<http://www.ine.es/prensa/np854.pdf>

[13]<http://www.ine.es/inebmenu/indice.htm>

[14]Chirivella González, V (2005) << Apuntes de econometría>> Departamento de Estadística e Investigación Operativa Aplicadas y Calidad de la Universidad Politécnica de Valencia, España

Listado de Figuras

| | |
|--|----|
| Figura III. 1 Ejemplo de matriz de correlación..... | 20 |
| Figura III. 2. Ejemplo matriz de correlación Inversa..... | 21 |
| Figura III. 3. Ejemplo de un gráfico de residuo para un modelo con problemas de heterocedasticidad. | 24 |
| Figura III. 4. Ejemplo de residuos anómalos. | 24 |
| Figura III. 5. Ejemplo falta de linealidad. | 25 |
| Figura III. 6. Gráfico del residuo de un ajuste sin contraste..... | 25 |
| Figura III. 7. Ejemplo de residuo con problemas de heterocedasticidad y falta de linealidad. | 26 |
| Figura III. 8. Ejemplo Autocorrelación primer orden Negativa. | 26 |
| Figura III. 9. Ejemplo Autocorrelación primer orden Positiva. | 26 |
| Figura III. 10. Representación de la varianza de la perturbación cuando se da el problema de Heterocedasticidad..... | 27 |
| Figura III. 11. Grafico problema de Heterocedasticidad. | 28 |
| Figura III. 12. Funciones de autocorrelación simples y parciales de los residuos del ajuste. | 30 |
| Figura III. 13 Grafico predicción puntual y predicción por intervalos de confianza..... | 32 |
| | |
| Figura IV. 1 Matriz de correlación (parte A) | 37 |
| Figura IV. 2 Matriz de Correlación (parte B) | 37 |
| Figura IV. 3 Matriz de Correlación Inversa (parte A)..... | 38 |
| Figura IV. 4 Matriz de Correlación Inversa (parte B)..... | 38 |
| Figura IV. 5 Matriz de Correlación Final..... | 39 |
| Figura IV. 6. Matriz de Correlación Inversa Final..... | 39 |
| Figura IV. 7 Gráfico de Residuos frente a Becas_CCAA..... | 42 |
| Figura IV. 8 Gráfico de Residuos frente Becas_Univ..... | 43 |
| Figura IV. 9 Gráfico de Residuos frente IPC..... | 43 |
| Figura IV. 10 Gráfico de Residuos frente DESEMP_ESO | 43 |
| Figura IV. 11 Gráfico de Residuos frente I+D+I | 44 |
| Figura IV. 12 Gráfico de Residuos frente a PARO..... | 44 |
| Figura IV. 13 Gráfico de residuos frente a la variable Becas_MICIN | 45 |
| Figura IV. 14 Gráfico de residuos frente a la variable IPI | 45 |
| Figura IV. 15 Grafico de residuos frente a la variable Flujo de Migración. | 45 |
| Figura IV. 16 Gráfico de residuos frente a número de fila. | 46 |

| | |
|--|----|
| Figura IV. 17 Gráfico de residuos frente a la variable $Becas_MICIN^4$ | 46 |
| Figura IV. 18 Histograma de Residuos | 48 |
| Figura IV. 19 Gráfico Función de autocorrelación simple. | 49 |
| Figura IV. 20 Gráfico Función de autocorrelación parcial. | 50 |
| Figura IV. 21 Histograma de residuos..... | 55 |
| Figura IV. 22 Histograma de Residuos | 57 |
| Figura IV. 23 Gráfico Función de autocorrelación simple del modelo libre de Heterocedasticidad y con distribución Normal de los residuos..... | 57 |
| Figura IV. 24 Gráfico Función de autocorrelación parcial del modelo libre de Heterocedasticidad y con distribución Normal de los residuos..... | 57 |
| Figura IV. 25 Tabla de modificación de los valores de las X para la obtención de predicciones. | 59 |
| Figura IV. 26 Resultado de las predicciones en función de las modificaciones de las X, para $Flujo_Migración/(Becas_Univ^{0.4})$ | 59 |
| Figura IV. 27 Representación del valor de las predicciones para el flujo de migración en función de la variable X modificada. | 60 |
| Figura IV. 28 Gráfico explicativo eliminación de la constante..... | 62 |
| Figura IV. 29 Gráfico de residuos vs variable I+D+I | 64 |
| Figura IV. 30 Gráfico de residuos frente al flujo de emigración. | 64 |
| Figura IV. 31 Gráfico de residuos frente a número de fila. | 64 |
| Figura IV. 32 Histograma de residuos del segundo modelo planteado. | 65 |
| Figura IV. 33 Histograma de residuos..... | 67 |
| Figura IV. 34 Función de autocorrelación simple (FAS) | 68 |
| Figura IV. 35 Función de autocorrelación Parcial (FAP)..... | 69 |

Listado de Tablas

| | |
|--|----|
| Tabla IV. 1. Matriz de datos inicial (parte A) | 35 |
| Tabla IV. 2. Matriz de datos inicial (parte B) | 36 |
| Tabla IV. 3. Resumen análisis regresión múltiple. | 41 |
| Tabla IV. 4. Resumen de análisis para el modelo de regresión. | 42 |
| Tabla IV. 5. Resumen análisis regresión múltiple elevando la variable Becas_MICIN^4. | 47 |
| Tabla IV. 6. Resumen análisis regresión del modelo elevando la variable Becas_MICIN^4. | 47 |
| Tabla IV. 7. Pruebas de Normalidad. | 48 |
| Tabla IV. 8. Resumen Regresión Múltiple para la comprobación de heterocedasticidad. | 49 |
| Tabla IV. 9. Puntos influyentes..... | 50 |
| Tabla IV. 10. Residuos Atípicos | 51 |
| Tabla IV. 11. Resumen regresión múltiple | 51 |
| Tabla IV. 12. Resumen regresión múltiple sin la variable PARO..... | 52 |
| Tabla IV. 13. Resumen regresión múltiple sin la variable IPC | 52 |
| Tabla IV. 14. Resumen regresión múltiple sin la variable IPI | 52 |
| Tabla IV. 15. Resumen de regresión múltiple con las variables significativas | 53 |
| Tabla IV. 16. Búsqueda del valor de “h” para la solución de heterocedasticidad. | 54 |
| Tabla IV. 17. Modelo de regresión libre de Heterocedasticidad eliminando la constante | 54 |
| Tabla IV. 18. Resumen regresión múltiple para la comprobación de la eliminación de Heterocedasticidad..... | 55 |
| Tabla IV. 19. Pruebas de normalidad de los residuos | 56 |
| Tabla IV. 20. Residuos atípicos del modelo libre de Heterocedasticidad | 56 |
| Tabla IV. 21. Residuos atípicos del modelo libre de Heterocedasticidad eliminando los datos de Murcia..... | 56 |
| Tabla IV. 22. Pruebas para la normalidad de los residuos | 56 |
| Tabla IV. 23. Resumen regresión múltiple del modelo definitivo | 58 |
| Tabla IV. 24. Resumen análisis regresión múltiple | 58 |
| Tabla IV. 25. Resumen análisis regresión múltiple | 62 |
| Tabla IV. 26. Resumen regresión múltiple | 62 |
| Tabla IV. 27. Resumen regresión múltiple de la variable I+D+I..... | 63 |
| Tabla IV. 28. Resumen regresión múltiple segundo modelo | 63 |
| Tabla IV. 29. Test para la normalidad de los residuos..... | 65 |
| Tabla IV. 30. Residuos atípicos segundo modelo propuesto..... | 66 |

| | |
|--|----|
| Tabla IV. 31 Resumen regresión múltiple | 66 |
| Tabla IV. 32 Resumen regresión múltiple | 67 |
| Tabla IV. 33 Pruebas de normalidad de los residuos | 67 |
| Tabla IV. 34 Resumen Regresión Múltiple para la comprobación de heterocedasticidad. | 68 |
| Tabla IV. 35 Residuos atípicos..... | 69 |
| Tabla IV. 36 Resultado de la predicción | 70 |
| Tabla IV. 37 Comparación de la predicción del primer modelo con el segundo modelo..... | 70 |