



UNIVERSIDAD
DE GRANADA



LA TASA DE DESEMPLEO EN PERSONAS MENORES DE 25 EN LOS AÑOS 2008-2020

Econometría II Grupo A. Trabajo modelo
econométrico

Abril 2022

Elena Gómez Pacheco
elenagomez@correo.ugr.es

ÍNDICE

1. Resumen
2. Motivación
3. Introducción
4. Datos
 - 4.1. Descripción de los datos
 - 4.2. Tabla de los datos
 - 4.3. Estadísticos descriptivos básicos y matriz de correlación
 - 4.4. Gráficos descriptivos
 - 4.4.1. Gráfico de series temporales
 - 4.4.2. Gráfico de dispersión (scatter)
5. Resultados
 - 5.1. Especificación del modelo
 - 5.2. Estimación del modelo
 - 5.2.1. Modelo lineal agrupado
 - 5.2.2. Modelo de efectos fijos
 - 5.2.3. Modelo de efectos aleatorios
 - 5.2.4. Modelo between
6. Selección del modelo
 - 6.1. Árbol de decisión
 - 6.2. Modelo seleccionado
 - 6.2.1. Cumplimiento de las hipótesis básicas
 - 6.2.2. Modelo final seleccionado
7. Conclusiones
8. Bibliografía
9. Anexo

1. RESUMEN

A fin de estudiar el desempleo en personas menores de 25 años se ha realizado un modelo econométrico, teniendo en cuenta las comunidades autónomas de España y el periodo de tiempo 2008-2020, que analiza la relación de variables como el abandono de la formación temprano, la tasa de paro según si se posee una titulación universitaria y/o de posgrado, el crecimiento del PIB y la tasa de desempleo para los mayores de 25 años, con el desempleo juvenil; con el objetivo de indagar empíricamente acerca de las posibilidades de empleabilidad en menores de 25 años. Los resultados más significativos indican que existe una relación negativa entre el desempleo juvenil y el abandono temprano de la formación y con la tasa de desempleo de larga duración para la población con estudios de posgrado. Además, indican una relación positiva con la tasa de desempleo en mayores de 25 años. Por último, se describen algunas de las posibles medidas para tratar de solucionar este problema.

2. MOTIVACIÓN

La motivación principal de este trabajo es indagar acerca del desempleo juvenil debido a que nuestra entrada en el mercado laboral se aproxima. Actualmente curso el final del grado y tenemos que decidir si continuar formándonos o si comenzar a buscar nuestra posición en este nuevo mundo. Para ello he realizado una investigación y un modelo con el objetivo de comprobar si las posibilidades de empleo a nuestra edad concuerdan realmente con los pensamientos de la sociedad o son totalmente diferentes.

3. INTRODUCCIÓN

El desempleo juvenil es un problema saliente desde hace muchos años. Las tasas de paro para los menores de 25 años son cada vez mayores a pesar de que la formación es más completa y exigente año tras año. Pero esto no es solamente un problema nacional, en la Unión Europea también hay países que sufren este hecho. En este trabajo nos centraremos en la situación de España, el país que más desempleo juvenil sufre de la UE, junto con Grecia e Italia. En el caso de España este problema es estructural, debido a la situación del mercado de trabajo y de la enseñanza, entre otros. Por ello, el objetivo de este trabajo es analizar el problema del desempleo en personas menores de 25 años a través de distintas variables que nos ayuden a realizar el análisis y tratar de encontrar una explicación empírica.

Basándonos en la literatura hemos escogido unas variables explicativas que se relacionen con esta problemática. Hay dos variables concretas que tienen relación con el mercado de trabajo y de la educación. Con respecto a este último, pensamos que uno de los aspectos más destacables como causa del problema es el abandono temprano de la educación. En tiempos pasados, muchos jóvenes se veían obligados a abandonar su formación para poder ayudar económicamente a sus familias mientras que hoy día los jóvenes, en su gran mayoría, no tienen ese tipo de problemas ya que desde el gobierno se han implantado medidas de ayuda para los problemas económicos, como pueden las becas o la educación básica obligatoria pública, aunque existen otros. Actualmente, el abandono temprano en España sigue existiendo y en una gran proporción en comparación en la UE. No obstante, ahora son muchos más los que pueden formarse para poder conseguir por un buen futuro gracias a las ayudas anteriormente mencionadas, esto deriva en otros problemas que conllevan el abandono de la formación temprana. En concreto, debido al interés por la adquisición de bienes materiales (Miñaca & Hervás, 2013, como se citó en Arredondo & Vizcaíno, 2020), para ello se deben introducir al mercado laboral, y la desmotivación (Arredondo & Vizcaíno, 2020).

Respecto al mercado de trabajo, se ha considerado relevante incluir el desempleo de los jóvenes que poseen una titulación universitaria y de posgrado, y el desempleo de larga duración de los mismos. La mayoría de los jóvenes españoles tienen un contrato de duración temporal, en concreto, el 4% de los asalariados escogieron de forma voluntaria este tipo de contrato (García, 2014). Esto puede deberse a las duras sanciones del despido imprecendente del contrato fijo o a que las empresas los utilizan como forma de comprobar el tipo de productividad, es decir, como una especie de filtro (PORTUGAL y VAREJÃO, 2009 como se citó en García, 2014).

Por tanto, surge la siguiente hipótesis. Se espera que el abandono temprano escolar influya de manera positiva en la tasa de desempleo de los menores de 25 años, es decir, cuanto más abandono más desempleo juvenil; y las variables relacionadas con el paro de la formación universitaria y de posgrado influyan también de una forma positiva.

4. DATOS

4.1. Descripción de los datos

En el trabajo hay 792 datos de panel en total, de los cuales hay una combinación de 11 de corte transversal y 13 datos temporales. Los datos corresponden a 11 provincias españolas que son: Andalucía, Canarias, Principado de Asturias, Castilla La Mancha, Cataluña, Comunidad Valenciana, Galicia, Comunidad de Madrid, Región de Murcia, País Vasco y Ceuta; y corresponden al período de años 2008-2020. Las unidades de medida de la gran parte de los datos son porcentajes y tasas. Todos los datos son procedentes de Instituto Nacional de Estadística (INE), excepto los datos de la tasa de paro relacionado con los estudios universitarios y de posgrado que son elaboración propia del EAE Business School a partir de los datos del INE.

4.2. Tabla de los datos

Nombre de las variables	Descripción de las variables	Tipo de variables	Unidad de medida	Signo esperado
tas_des	Tasa desempleo menores de 25 años por comunidad autónoma.	Explicada/ cuantitativa continua	Tasas	
tas_paro	Tasas de paro mayores de 25 años por comunidad autónoma.	Explicativa/ cuantitativa continua	Tasas	+
aband_temp	Abandono temprano de la educación-formación en la población de 18 a 24 años.	Explicativa/ cuantitativa continua	Porcentaje	+
crec_pib	Tasa de crecimiento anual del PIB real per cápita.	Explicativa/ cuantitativa continua	Porcentaje	+

tasparo_uni	Evolución de la tasa de paro para la población con estudios universitarios.	Explicativa/cuantitativa continua	Porcentaje	+
tasparo_pos	Evolución de la tasa de paro para la población con estudios de posgrado.	Explicativa/cuantitativa continua	Porcentaje	+
tasuni_larg	Evolución de la tasa de paro de larga duración para la población con estudios universitarios.	Explicativa/cuantitativa continua	Porcentaje	+
taspos_larg	Evolución de la tasa de paro de larga duración para la población con estudios de posgrado.	Explicativa/cuantitativa continua	Porcentaje	+

4.3. Estadísticos descriptivos básicos y matriz de correlación

Estadísticos principales, usando las observaciones 1:01 - 11:13

(se ignoraron los valores ausentes)

Variable	Media	Mediana	D. T.	Mín	Máx
tas_des	43.5	42.9	12.0	19.5	79.4
aband_temp	22.6	21.9	7.99	6.50	45.8
tas_paro	18.0	16.9	6.34	5.59	33.4
crec_pib	0.369	1.50	4.06	-17.5	6.70
tasparo_uni	11.1	10.6	3.32	6.10	20.4
tasparo_pos	6.18	5.60	2.68	2.00	12.1
tasuni_larg	56.8	56.8	7.13	42.2	69.8

taspos_larg	46.2	48.2	10.4	22.6	63.9
-------------	------	------	------	------	------

MATRIZ DE CORRELACIÓN

Coefficientes de correlación, usando las observaciones 1:02 - 10:06

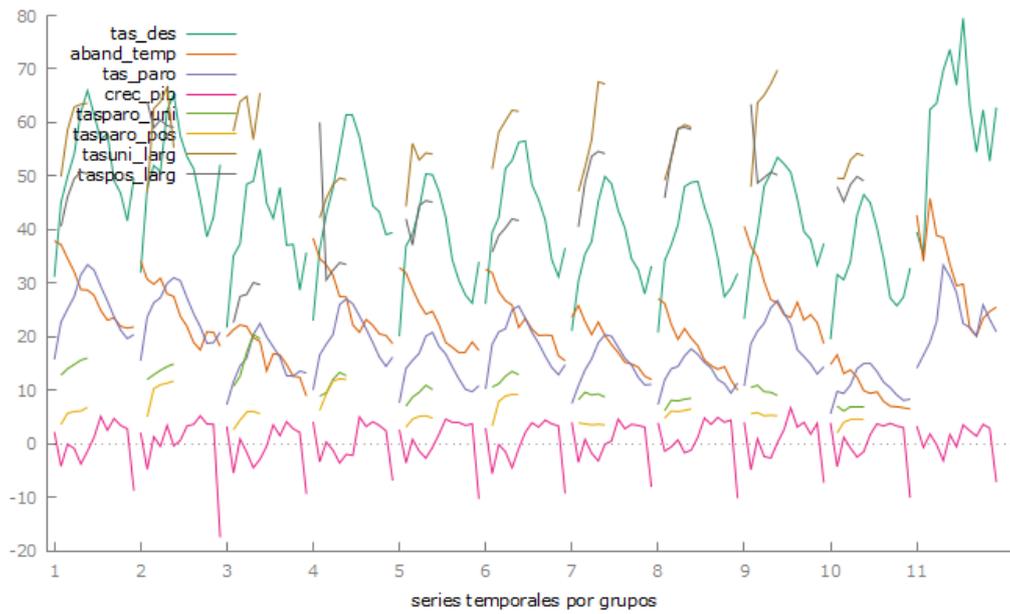
(se ignoraron los valores ausentes)

Valor crítico al 5% (a dos colas) = 0.1779 para n = 122

tas_des	aband_temp	tas_paro	crec_pib	tasparo_uni	
1.0000	0.1847	0.9291	-0.1395	0.6570	tas_des
	1.0000	0.3754	-0.0306	0.2718	aband_temp
		1.0000	-0.0804	0.6645	tas_paro
			1.0000	-0.0955	crec_pib
				1.0000	tasparo_uni
		tasparo_pos	tasuni_larg	taspos_larg	
		0.6247	0.4959	0.0901	tas_des
		0.2930	-0.1258	0.0645	aband_temp
		0.6179	0.4542	0.1885	tas_paro
		0.2947	0.3151	0.1827	crec_pib
		0.4396	0.3789	-0.2874	tasparo_uni
		1.0000	0.0773	0.0204	tasparo_pos
			1.0000	0.0901	tasuni_larg
				1.0000	taspos_larg

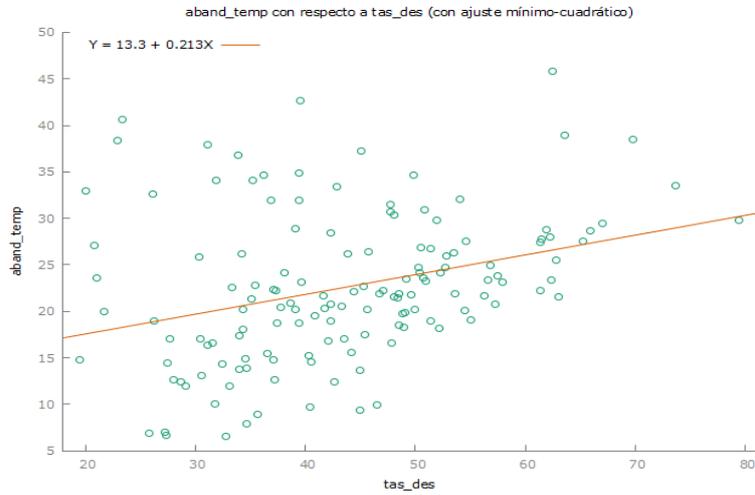
4.4. Gráficos descriptivos

4.4.1. Gráfico de series temporales

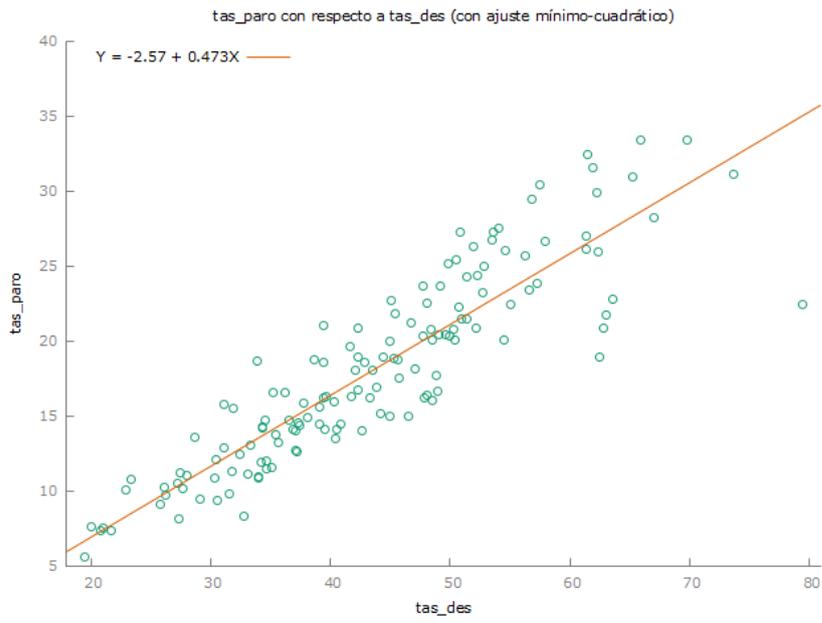


4.4.2. Gráfico de dispersión (scatter):

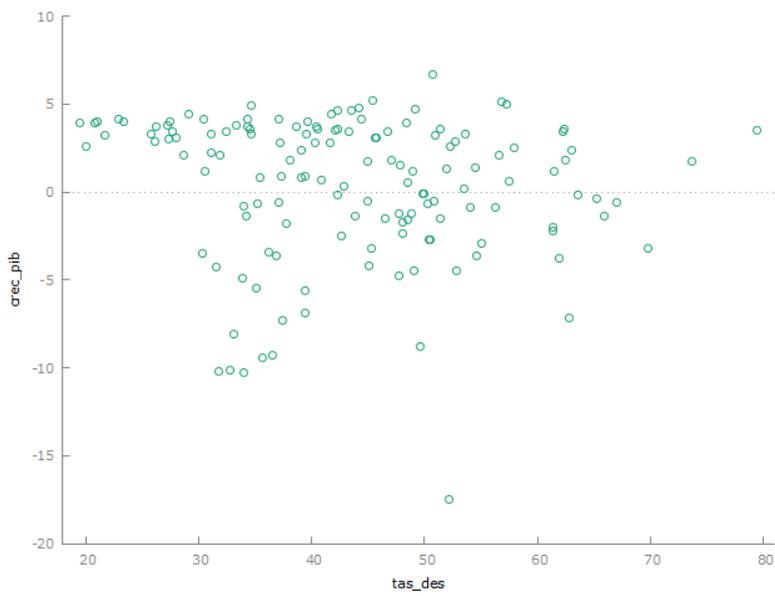
- Relación tas_des con aband_temp:



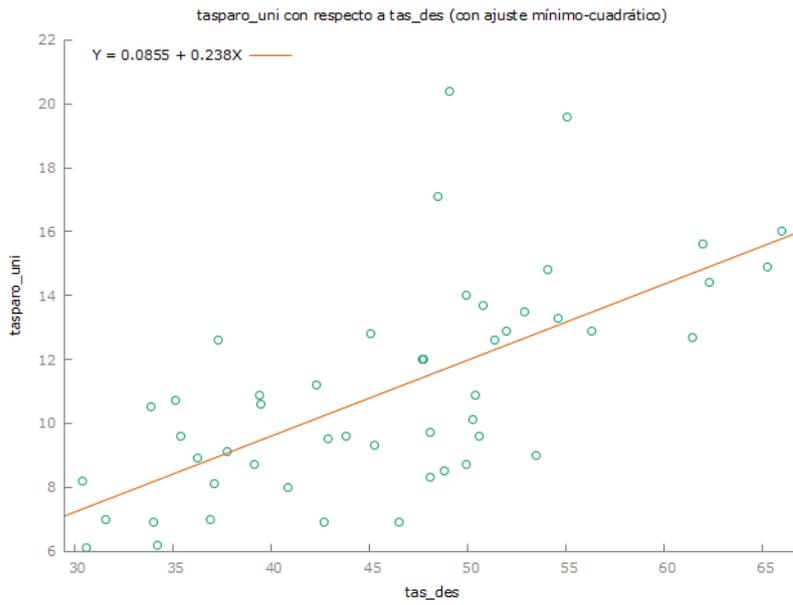
- Relación tas_des con tas_paro:



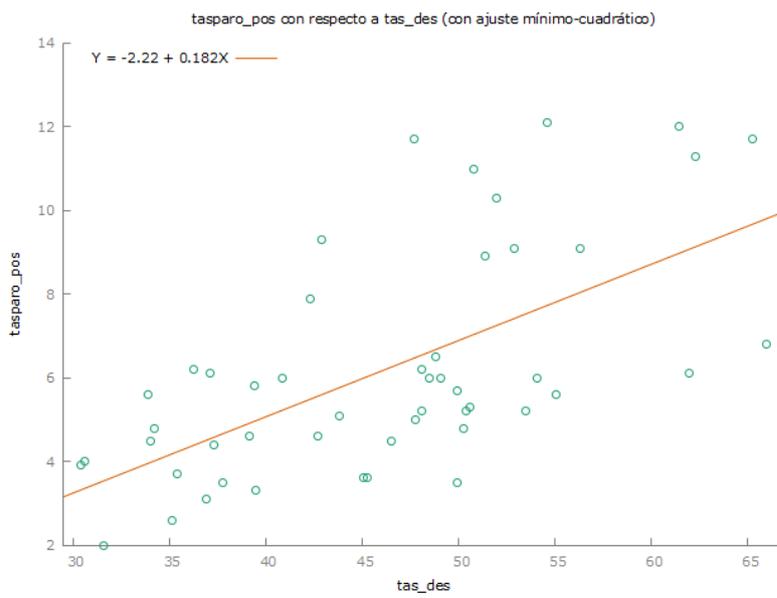
- Relación tas_des con crec_pib:



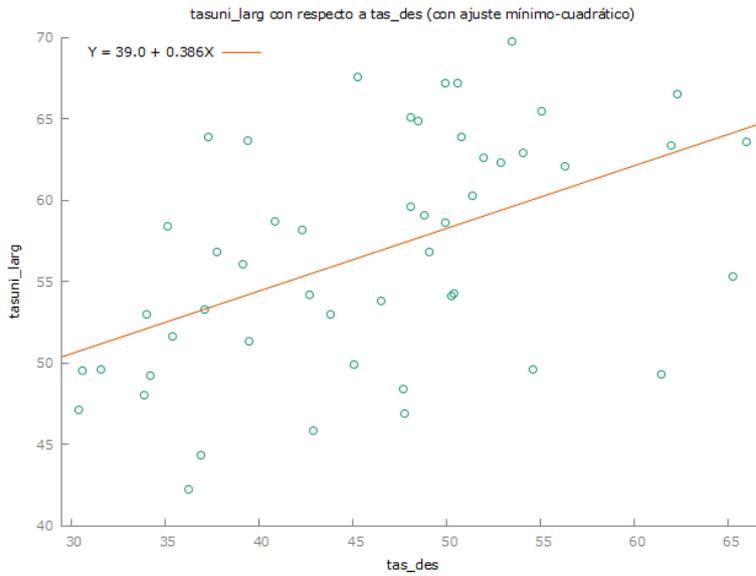
- Relación tas_des con tasparo_uni:



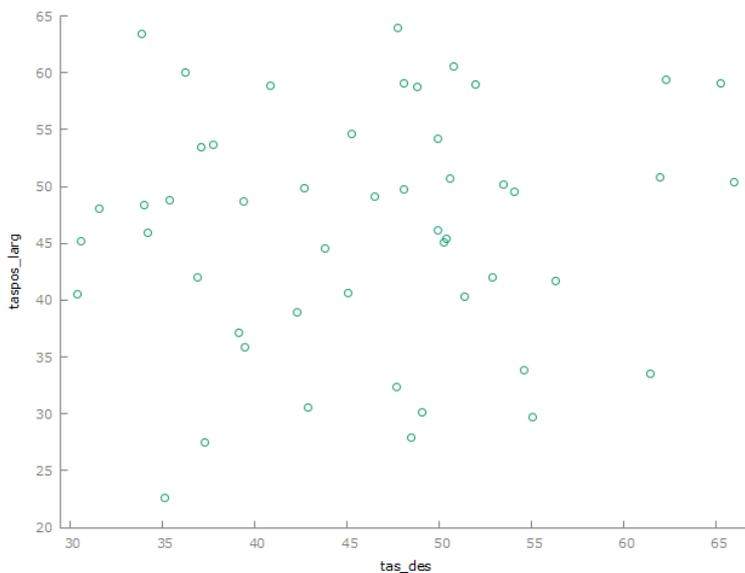
- Relación tas_des con tasparo_pos:



- Relación tas_des con tasuni_larg:



- Relación tas_des con taspos_larg:



Como podemos observar en los gráficos, la variable que más correlación tiene es la tasa de desempleo en personas mayores de 25 años, seguida del porcentaje de abandono temprano de la formación y por último las tasas de paro para la población con estudios universitarios y de posgrado. La tasa de paro de larga duración para la población con estudios universitarios también tiene correlación pero muy débil.

5. RESULTADOS

5.1. Especificación del modelo

$$\begin{aligned}tas_des = & \alpha_{it} + \beta_1 aband_temp_{it} + \beta_2 tas_paro_{it} + \beta_3 crec_pib_{it} \\ & + \beta_4 tasparo_uni_{it} + \beta_5 tasparo_pos_{it} + \beta_6 tasuni_larg_{it} \\ & + \beta_7 taspos_larg_{it} + u_{it}\end{aligned}$$

La i representa las comunidades autónomas $i= 1, \dots, 11$.

La t representa el periodo de tiempo $t= 1, \dots, 13$.

Dependiendo del valor del término α , vamos a distinguir tres tipos de modelo (Chica & Salmerón, 2019):

- Modelo agrupado: el término α_{it} es constante para todas las comunidades autónomas y en el periodo 2008-2020. ($\alpha_{it} = \alpha$)
- Modelo de efectos fijos: el término α_{it} es diferente para cada comunidad autónoma, cada periodo o para ambos.
- Modelo de efectos aleatorios: en este caso α va a ser una variable aleatoria.

5.2. Estimación del modelo

5.2.1. MODELO LINEAL AGRUPADO

Modelo 4: MCO combinados, utilizando 50 observaciones

Se han incluido 10 unidades de sección cruzada

Largura de la serie temporal = 5

Variable dependiente: tas_des

	<i>Coficiente</i>	<i>Desv. Típica</i>	<i>Estadístico t</i>	<i>valor p</i>	
const	30.9444	5.60408	5.522	<0.0001	***
aband_temp	-0.515861	0.0806547	-6.396	<0.0001	***
tas_paro	1.57087	0.146324	10.74	<0.0001	***

crec_pib	-0.00490449	0.245159	-0.02001	0.9841
tasparo_uni	0.0313288	0.201095	0.1558	0.8769
tasparo_pos	0.287608	0.230619	1.247	0.2193
tasuni_larg	-0.0339289	0.0838085	-0.4048	0.6877
taspos_larg	-0.0699072	0.0477675	-1.463	0.1508
Media de la vble. dep.	46.08760	D.T. de la vble. dep.	9.168398	
Suma de cuad. residuos	324.6806	D.T. de la regresión	2.780376	
R-cuadrado	0.921173	R-cuadrado corregido	0.908036	
F(7, 42)	70.11634	Valor p (de F)	4.13e-21	
Log-verosimilitud	-117.7174	Criterio de Akaike	251.4348	
Criterio de Schwarz	266.7310	Crit. de Hannan-Quinn	257.2597	
rho	0.249383	Durbin-Watson	1.320878	

Contraste de normalidad de los residuos -

Hipótesis nula: [El error tiene distribución Normal]

Estadístico de contraste: Chi-cuadrado(2) = 0.593524

con valor p = 0.743221

5.2.2. EFECTOS FIJOS:

Modelo 4: Efectos fijos, utilizando 50 observaciones

Se han incluido 10 unidades de sección cruzada

Largura de la serie temporal = 5

Variable dependiente: tas_des

	<i>Coefficiente</i>	<i>Desv. Típica</i>	<i>Estadístico t</i>	<i>valor p</i>	
const	17.1366	10.9010	1.572	0.1255	
aband_temp	-0.359190	0.245414	-1.464	0.1528	
tas_paro	1.76910	0.307053	5.762	<0.0001	***
crec_pib	-0.180180	0.240852	-0.7481	0.4597	
tasparo_uni	-0.0353349	0.398406	-0.08869	0.9299	
tasparo_pos	0.213155	0.405691	0.5254	0.6028	
tasuni_larg	0.0722021	0.112027	0.6445	0.5237	
taspos_larg	-0.0544002	0.0726779	-0.7485	0.4595	

Media de la vble. dep.	46.08760	D.T. de la vble. dep.	9.168398
------------------------	----------	-----------------------	----------

Suma de cuad. residuos	178.2157	D.T. de la regresión	2.323893
R-cuadrado MCVF (LSDV)	0.956732	R-cuadrado 'intra'	0.927107
F(16, 33) MCVF	45.60593	Valor p (de F)	6.85e-18
Log-verosimilitud	-102.7212	Criterio de Akaike	239.4424
Criterio de Schwarz	271.9468	Crit. de Hannan-Quinn	251.8203
rho	-0.288642	Durbin-Watson	2.322883

Contraste conjunto de los regresores (excepto la constante) -

Estadístico de contraste: $F(7, 33) = 59.9602$

con valor $p = P(F(7, 33) > 59.9602) = 6.13539e-17$

Contraste de diferentes interceptos por grupos -

Hipótesis nula: [Los grupos tienen un intercepto común]

Estadístico de contraste: $F(9, 33) = 3.01341$

con valor $p = P(F(9, 33) > 3.01341) = 0.00975066$

5.2.3. EFECTOS ALEATORIOS:

Modelo 5: Efectos aleatorios (MCG), utilizando 50 observaciones

Se han incluido 10 unidades de sección cruzada

Largura de la serie temporal = 5

Variable dependiente: *tas_des*

	<i>Coefficiente</i>	<i>Desv. Típica</i>	<i>z</i>	<i>valor p</i>	
const	27.5123	6.10614	4.506	<0.0001	***
aband_temp	-0.530605	0.0882266	-6.014	<0.0001	***
tas_paro	1.60023	0.151784	10.54	<0.0001	***
crec_pib	-0.141202	0.225110	-0.6273	0.5305	
tasparο_uni	-0.0159081	0.220241	-0.07223	0.9424	
tasparο_pos	0.311096	0.263416	1.181	0.2376	
tasuni_larg	0.0203113	0.0869438	0.2336	0.8153	
taspos_larg	-0.0635952	0.0513918	-1.237	0.2159	

Media de la vble. dep.	46.08760	D.T. de la vble. dep.	9.168398
Suma de cuad. residuos	334.1858	D.T. de la regresión	2.787788
Log-verosimilitud	-118.4388	Criterio de Akaike	252.8776
Criterio de Schwarz	268.1737	Crit. de Hannan-Quinn	258.7024
rho	-0.288642	Durbin-Watson	2.322883

Varianza 'entre' (between) = 2.30237

Varianza 'dentro' (Within) = 5.40048

theta usado para quasi-demeaning (cuasi-centrado de los datos) = 0.434914

Contraste conjunto de los regresores (excepto la constante) -

Estadístico de contraste asintótico: Chi-cuadrado(7) = 495.24

con valor $p = 8.45856e-103$

Contraste de Breusch-Pagan -

Hipótesis nula: [Varianza del error específico a la unidad = 0]

Estadístico de contraste asintótico: Chi-cuadrado(1) = 5.18448

con valor $p = 0.0227894$

Contraste de Hausman -

Hipótesis nula: [Los estimadores de MCG son consistentes]

Estadístico de contraste asintótico: Chi-cuadrado(7) = 11.6152

con valor $p = 0.113943$

El contraste conjunto de los regresores nos indica que el modelo es significativo globalmente, debido a que el estadístico es un valor muy grande y será mayor que el que aparece en las tablas.

5.2.4. MODELO BETWEEN:

En el modelo “entre” o “between” se utilizan las medias de los datos temporales de cada individuo. En la práctica se suelen utilizar pocas veces debido a que ignoran la información temporal que existe dentro de los individuos (Chica & Salmerón, 2019).

Modelo 6: Entre grupos, usando las observaciones 1-10

Variable dependiente: tas_des

	<i>Coefficiente</i>	<i>Desv. Típica</i>	<i>Estadístico t</i>	<i>valor p</i>
const	37.0909	20.3617	1.822	0.2101
aband_temp	-0.0811044	0.319943	-0.2535	0.8236
tas_paro	0.942802	0.765450	1.232	0.3432
crec_pib	1.32394	3.53984	0.3740	0.7443
tasparo_uni	0.756173	1.08096	0.6995	0.5566
tasparo_pos	-0.0322801	0.484720	-0.06660	0.9530
tasuni_larg	-0.228179	0.297328	-0.7674	0.5230
taspos_larg	-0.0158815	0.447015	-0.03553	0.9749
Media de la vble. dep.	46.08760	D.T. de la vble. dep.	6.099189	
Suma de cuad. residuos	6.764933	D.T. de la regresión	1.839148	
R-cuadrado	0.979794	R-cuadrado corregido	0.909074	
F(7, 2)	13.85447	Valor p (de F)	0.068952	
Log-verosimilitud	-12.23522	Criterio de Akaike	40.47044	
Criterio de Schwarz	42.89112	Crit. de Hannan-Quinn	37.81496	

6. SELECCIÓN DE MODELO

Una vez ya hemos obtenido los tres modelos, el agrupado, el de efectos fijos y el de efectos aleatorios; debemos seleccionar el que nos parezca más adecuado y para ello usaremos tres contrastes diferentes.

TEST DE CHOW

El Test de Chow es un contraste de efectos fijos y sirve para contrastar si es más conveniente un modelo con datos agrupados, es decir, pooled o un modelo de efectos fijos.

Para ello realizamos dos hipótesis, la nula y la alternativa. En la nula tenemos que es más adecuado un modelo agrupado y en la hipótesis alternativa el modelo más acertado es el de efectos fijos.

H_0 : Modelo agrupado

H_1 : Modelo efectos fijos

A continuación, los resultados del Test de Chow son:

Regresión aumentada para el contraste de Chow

MCO, utilizando 50 observaciones

Variable dependiente: tas_des

	coeficiente	Desv. típica	Estadístico t	valor p	
const	50.5485	9.91496	5.098	1.28e-05	***
aband_temp	-0.699076	0.162175	-4.311	0.0001	***
tas_paro	1.69963	0.178971	9.497	4.30e-011	***
crec_pib	0.547632	0.382740	1.431	0.1616	
tasparo_uni	0.0351489	0.293149	0.1199	0.9053	
tasparo_pos	-0.125406	0.270356	-0.4639	0.6457	
tasuni_larg	-0.250813	0.142733	-1.757	0.0879	*
taspos_larg	-0.0874278	0.0600096	-1.457	0.1543	
splitdum	-39.9322	13.0324	-3.064	0.0043	***
sd_aband_temp	0.220548	0.195175	1.130	0.2664	
sd_tas_paro	-0.575349	0.352963	-1.630	0.1123	
sd_crec_pib	-1.14813	0.505517	-2.271	0.0296	**
sd_tasparo_uni	-0.444385	0.633188	-0.7018	0.4876	
sd_tasparo_pos	1.25222	0.511230	2.449	0.0196	**
sd_tasuni_larg	0.581608	0.226494	2.568	0.0148	**
sd_taspos_larg	0.104440	0.103140	1.013	0.3184	
Media de la vble. dep.	46.08760	D.T. de la vble. dep.	9.168398		
Suma de cuad. residuos	206.5523	D.T. de la regresión	2.464765		
R-cuadrado	0.949853	R-cuadrado corregido	0.927729		
F(15, 34)	42.93357	Valor p (de F)	1.11e-17		
Log-verosimilitud	-106.4102	Criterio de Akaike	244.8204		
Criterio de Schwarz	275.4127	Crit. de Hannan-Quinn	256.4701		

Contraste de Chow de cambio estructural en la observación 6:02

$F(8, 34) = 2.4306$ con valor p 0.0339

Como resultados obtenemos que el p-valor es menor que el nivel de significación (5%), por lo tanto se rechaza la hipótesis nula. El modelo más adecuado es el de efectos fijos.

TEST DE BREUSCH-PAGAN

El contraste de Breusch-Pagan trata de identificar que modelo sería el más adecuado, el modelo pooled o el de efectos aleatorios. Para ello realizamos una hipótesis nula, que sería que la varianza que corresponde a los individuos es cero, mientras que la hipótesis alternativa es lo contrario, que la varianza correspondiente a los individuos sea distinta de cero.

H_0 : Modelo agrupado (pooled)

H_1 : Modelo efectos aleatorios

Contraste de Breusch-Pagan -

Hipótesis nula: [Varianza del error específico a la unidad = 0]

Estadístico de contraste asintótico: Chi-cuadrado(1) = 5.18448

con valor $p = 0.0227894$

En el resultado observamos que el p-valor es menor que el nivel de significación (5%), por lo tanto se rechazaría la hipótesis nula y el modelo más adecuado es el de efectos fijos.

TEST DE HAUSMAN

Con el resultado de los dos test anteriores el modelo agrupado queda descartado, por tanto el test de Hausman contrasta si es más adecuado un modelo de efectos fijos o de efectos aleatorios. Por lo tanto, realizamos un contraste en el que la hipótesis nula es el del modelo de efectos aleatorios, mientras que la hipótesis alternativa es el modelo de efectos fijos.

H_0 : Modelo de efectos aleatorios

H_1 : Modelo efectos fijos

Contraste de Hausman -

Hipótesis nula: [Los estimadores de MCG son consistentes]

Estadístico de contraste asintótico: Chi-cuadrado(7) = 11.6152

con valor $p = 0.113943$

Como podemos observar en el resultado, el p-valor es mayor que el nivel de significación (5%), por tanto no podemos rechazar la hipótesis nula. Por consiguiente, el modelo más adecuado es el de efectos aleatorios.

6.1. Árbol de decisión

Test de Chow:

H_0 : Modelo agrupado → Se rechaza

H_1 : Modelo efectos fijos

Test de Breusch-Pagan:

H_0 : Modelo pooled → Se rechaza

H_1 : Modelo efectos aleatorios

Test de Hausman:

H_0 : Modelo de efectos aleatorios

H_1 : Modelo efectos fijos → Se rechaza

6.2. Modelo seleccionado

El modelo que se ha seleccionado es el de efectos aleatorios, debido a que ha salido como resultado en los test de contrastes y porque es el modelo con más variables significativas. Ahora comprobaremos que cumple todas las hipótesis básicas del modelo.

6.2.1. Cumplimiento de las hipótesis básicas del modelo

- **Colinealidad**

7. Diagnósticos de colinealidad de Belsley-Kuh-Welsch:

8.

9. proporciones de la varianza

10.

	lambda	cond	const	aband_te~	tas_paro	crec_pib	tasparo_~	tasparo_~
11.	tasuni_l~							
12.	6.721	1.000	0.000	0.001	0.001	0.002	0.001	0.002
13.	0.857	2.800	0.000	0.000	0.000	0.505	0.000	0.008
14.	0.191	5.934	0.004	0.036	0.012	0.097	0.029	0.188
15.	0.085	8.902	0.000	0.412	0.040	0.000	0.019	0.153
16.	0.071	9.739	0.000	0.067	0.001	0.088	0.245	0.364
17.	0.041	12.776	0.048	0.132	0.352	0.033	0.009	0.135
18.	0.029	15.232	0.004	0.010	0.191	0.061	0.674	0.000
19.	0.005	35.122	0.944	0.342	0.402	0.212	0.023	0.150

	lambda	cond	taspos_l~
20.			
21.	tasuni_l~		
22.	6.721	1.000	0.001
23.	0.857	2.800	0.000
24.	0.191	5.934	0.061
25.	0.085	8.902	0.044
26.	0.071	9.739	0.242
27.	0.041	12.776	0.074
28.	0.029	15.232	0.455
29.	0.005	35.122	0.123

- 30.
31. lambda = autovalores de la inversa de la matriz de covarianzas (smallest is 0.00544856)
32. cond = índice de condición
33. nota: Las columnas de proporciones de la varianza suman 1.0
- 34.
35. De acuerdo con BKW, cond >= 30 indica "fuerte" dependencia casi lineal,
36. y cond entre 10 y 30 "moderadamente fuerte". Las estimaciones de los parámetros cuya
37. varianza está principalmente asociada con valores cond problemáticos pueden así mismo
38. considerarse problemáticas.
- 39.
40. Cuenta de índices de condición >= 30: 1
41. Proporciones de varianza >= 0.5 asociadas con condición >= 30:
- 42.

```

43.      const tasuni_l~
44.      0.944      0.829
45.
46. Cuenta de índices de condición >= 10: 3
47. Proporciones de varianza >= 0.5 asociadas con condición >= 10:
48.
49.      const  tas_paro  tasparo_~  tasuni_l~  taspos_l~
50.      0.996      0.945      0.706      0.975      0.653

```

El Factor Inflación Varianza (FIV) nos permite poder cuantificar la intensidad de la colinealidad. En nuestro caso podemos observar que ningún valor supera el 10, por lo tanto no presenta colinealidad según el FIV.

Otro indicador es el test de Belssey, Kuh y Welsh, el cual nos permite conocer si existe una colinealidad elevada. Para ello se debe realizar un índice de condición, que se obtiene realizando la raíz cuadrada del auto-valor más grande de la matriz varianzas-covarianzas entre el valor propio; si el resultado se encuentra entre 10 y 30 la colinealidad no se considera lo suficientemente grande como para dar problemas y si es mayor que 30 el modelo presenta colinealidad elevada.

En nuestro caso observamos que ningún valor supera ni siquiera el valor 10, por tanto según este test tampoco presenta colinealidad nuestro modelo.

- Normalidad de los residuos

Para contrastar si los residuos siguen una normalidad en su distribución realizaremos un contraste, en el cual la hipótesis nula sería la normalidad de los residuos y la hipótesis alternativa sería lo contrario, los residuos no siguen una distribución normal.

H₀: Normalidad de los residuos

H₁: No normalidad de los residuos

Distribución de frecuencias para residual, observaciones 2-123
 número de cajas = 7, Media = 2.77112e-015, Desv.típ.=2.82078

intervalo	punto medio	frecuencia	rel	acum.
-----------	-------------	------------	-----	-------

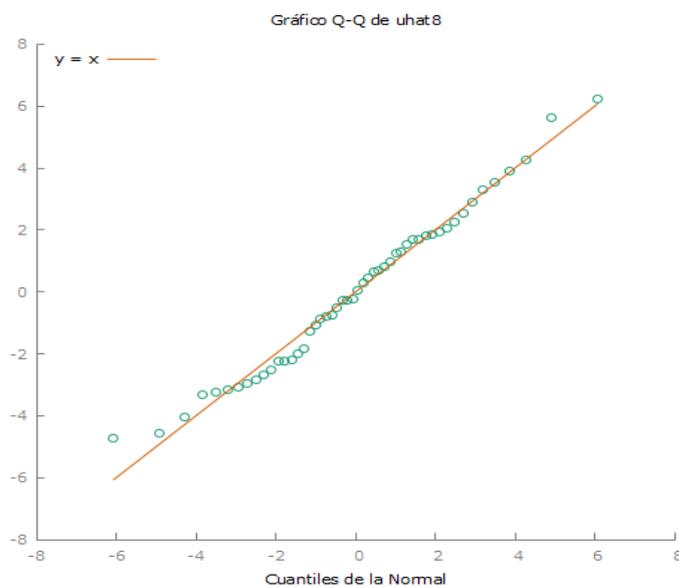
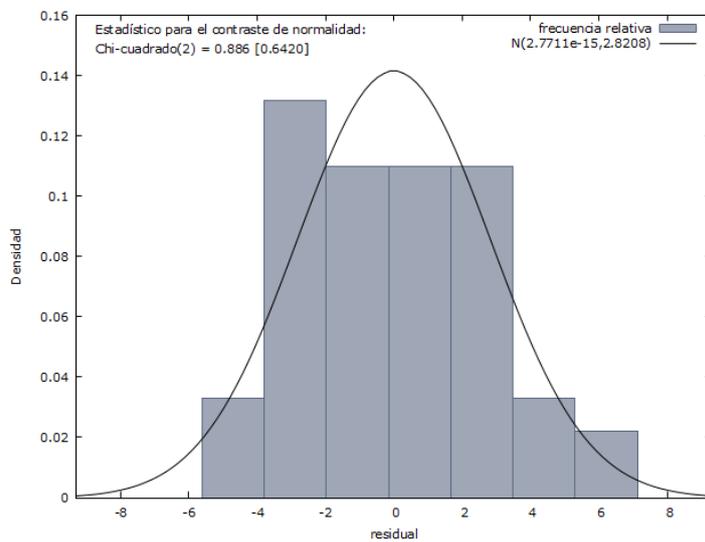
< -3.8002	-4.7109	3	6.00%	6.00%	**
-3.8002 -	-1.9787	12	24.00%	30.00%	*****
-1.9787 -	-0.15716	10	20.00%	50.00%	*****
-0.15716 -	1.6643	10	20.00%	70.00%	*****
1.6643 -	3.4858	10	20.00%	90.00%	*****
3.4858 -	5.3074	3	6.00%	96.00%	**
>= 5.3074	6.2181	2	4.00%	100.00%	*

Observaciones ausentes = 72 (59.02%)

Contraste de la hipótesis nula de distribución Normal:

Chi-cuadrado(2) = 0.886 con valor p 0.64202

Histograma:



En nuestro caso el p-valor es mayor que el nivel de significación (5%), por lo tanto no rechazamos la hipótesis nula, es decir, los residuos siguen una distribución normal.

- **Autocorrelación**

La autocorrelación significa que las covarianzas de las perturbaciones son no nulas para instantes de tiempo distintos. Para saber si autocorrelación en el modelo planteamos una hipótesis nula, que el modelo es incorrelado, y una hipótesis alternativa, que el modelo es correlado.

H_0 : Modelo incorrelado

H_1 : Modelo correlado

Ecuación en primeras diferencias (dependiente, d_y):

	coeficiente	Desv. típica	Estadístico t	valor p	
d_aband_temp	-0.504514	0.316042	-1.596	0.1449	
d_tas_paro	1.27531	0.577881	2.207	0.0547	*
d_crec_pib	-0.235653	0.256677	-0.9181	0.3825	
d_tasparo_uni	0.355529	0.560080	0.6348	0.5414	
d_tasparo_pos	-0.143773	0.343019	-0.4191	0.6849	
d_tasuni_larg	0.178805	0.179112	0.9983	0.3442	
d_taspos_larg	-0.107451	0.0367323	-2.925	0.0169	**

n = 40, R-squared = 0.6741

Autorregresión de los residuos (dependiente, \hat{u}):

	coeficiente	Desv. típica	Estadístico t	valor p	
$\hat{u}(-1)$	-0.457796	0.143925	-3.181	0.0112	**

n = 30, R-squared = 0.1838

Contraste de Wooldridge de autocorrelación en datos de panel -

Hipótesis nula: No autocorrelación de primer orden ($\rho = -0.5$)

Estadístico de contraste: $F(1, 9) = 0.0859881$

con valor p = $P(F(1, 9) > 0.0859881) = 0.77599$

Podemos observar en el contraste de Wooldridge que como el p-valor es mayor que el nivel de significación (5%), no podemos rechazar la hipótesis nula. Nuestro modelo seleccionado es incorrelado, es decir, las covarianzas de las perturbaciones aleatorias son nulas.

- **Heterocedasticidad**

Para comprobar si existe heterocedasticidad en el modelo vamos a comparar el modelo de efectos aleatorios y el modelo de efectos aleatorios con estimaciones robustas, si las desviaciones típicas son las mismas podemos decir que el modelo no presenta heterocedasticidad (Pérez & López, 2020, pág. 221).

- Modelo efectos aleatorios.

Modelo 5: Efectos aleatorios (MCG), utilizando 50 observaciones
 Se han incluido 10 unidades de sección cruzada
 Largura de la serie temporal = 5
 Variable dependiente: tas_des

	<i>Coefficiente</i>	<i>Desv. Típica</i>	<i>z</i>	<i>valor p</i>	
const	27.5123	6.10614	4.506	<0.0001	***
aband_temp	-0.530605	0.0882266	-6.014	<0.0001	***
tas_paro	1.60023	0.151784	10.54	<0.0001	***
crec_pib	-0.141202	0.225110	-0.6273	0.5305	
tasparo_uni	-0.0159081	0.220241	-0.07223	0.9424	
tasparo_pos	0.311096	0.263416	1.181	0.2376	
tasuni_larg	0.0203113	0.0869438	0.2336	0.8153	
taspos_larg	-0.0635952	0.0513918	-1.237	0.2159	
-					
Media de la vble. dep.	46.08760	D.T. de la vble. dep.	9.168398		
Suma de cuad. residuos	334.1858	D.T. de la regresión	2.787788		
Log-verosimilitud	-118.4388	Criterio de Akaike	252.8776		
Criterio de Schwarz	268.1737	Crit. de Hannan-Quinn	258.7024		
rho	-0.288642	Durbin-Watson	2.322883		

Varianza 'entre' (between) = 2.30237

Varianza 'dentro' (Within) = 5.40048

theta usado para quasi-demeaning (cuasi-centrado de los datos) = 0.434914
 Contraste conjunto de los regresores (excepto la constante) -
 Estadístico de contraste asintótico: Chi-cuadrado(7) = 495.24
 con valor p = 8.45856e-103

Contraste de Breusch-Pagan -
 Hipótesis nula: [Varianza del error específico a la unidad = 0]
 Estadístico de contraste asintótico: Chi-cuadrado(1) = 5.18448
 con valor p = 0.0227894

Contraste de Hausman -
 Hipótesis nula: [Los estimadores de MCG son consistentes]
 Estadístico de contraste asintótico: Chi-cuadrado(7) = 11.6152
 con valor p = 0.113943

- Modelo efectos aleatorios con desviaciones robustas (HAC)

Modelo 6: Efectos aleatorios (MCG), utilizando 50 observaciones

Se han incluido 10 unidades de sección cruzada

Largura de la serie temporal = 5

Variable dependiente: tas_des

Desviaciones típicas robustas (HAC)

	<i>Coficiente</i>	<i>Desv. Típica</i>	<i>z</i>	<i>valor p</i>	
const	27.5123	8.44523	3.258	0.0011	***
aband_temp	-0.530605	0.0836726	-6.341	<0.0001	***
tas_paro	1.60023	0.181119	8.835	<0.0001	***
crec_pib	-0.141202	0.207149	-0.6816	0.4955	
tasparo_uni	-0.0159081	0.154341	-0.1031	0.9179	
tasparo_pos	0.311096	0.240077	1.296	0.1950	
tasuni_larg	0.0203113	0.112549	0.1805	0.8568	
taspos_larg	-0.0635952	0.0442286	-1.438	0.1505	

Media de la vble. dep.	46.08760	D.T. de la vble. dep.	9.168398
Suma de cuad. residuos	334.1858	D.T. de la regresión	2.787788
Log-verosimilitud	-118.4388	Criterio de Akaike	252.8776
Criterio de Schwarz	268.1737	Crit. de Hannan-Quinn	258.7024
rho	-0.288642	Durbin-Watson	2.322883

Varianza 'entre' (between) = 2.30237

Varianza 'dentro' (Within) = 5.40048

theta usado para quasi-demeaning (cuasi-centrado de los datos) = 0.434914

Contraste conjunto de los regresores (excepto la constante) -

Estadístico de contraste asintótico: Chi-cuadrado(7) = 1705.11

con valor $p = 0$

Contraste de Breusch-Pagan -

Hipótesis nula: [Varianza del error específico a la unidad = 0]

Estadístico de contraste asintótico: Chi-cuadrado(1) = 5.18448

con valor $p = 0.0227894$

Contraste de Hausman -

Hipótesis nula: [Los estimadores de MCG son consistentes]

Estadístico de contraste asintótico: Chi-cuadrado(7) = 23.8223

con valor $p = 0.00122424$

Como podemos observar, las desviaciones típicas son diferentes. Por lo tanto, nuestro modelo presenta heterocedasticidad.

Para solucionarlo utilizaremos el Estimador Mínimos Cuadrados Generalizados Factibles.

Modelo 7: Con corrección de heterocedasticidad, utilizando 50 observaciones

Variable dependiente: *tas_des*

	<i>Coefficiente</i>	<i>Desv. Típica</i>	<i>Estadístico t</i>	<i>valor p</i>	
const	38.3302	4.20351	9.119	<0.0001	***
aband_temp	-0.560577	0.0468649	-11.96	<0.0001	***
tas_paro	1.74399	0.0908524	19.20	<0.0001	***
crec_pib	0.165792	0.177221	0.9355	0.3549	
tasparo_uni	-0.00712281	0.157487	-0.04523	0.9641	
tasparo_pos	0.142435	0.183958	0.7743	0.4431	
tasuni_larg	-0.141904	0.0628859	-2.257	0.0293	**
taspos_larg	-0.121699	0.0348602	-3.491	0.0011	***

Estadísticos basados en los datos ponderados:

Suma de cuad. residuos	106.8498	D.T. de la regresión	1.595006
R-cuadrado	0.980950	R-cuadrado corregido	0.977775
F(7, 42)	308.9594	Valor p (de F)	5.35e-34
Log-verosimilitud	-89.93196	Criterio de Akaike	195.8639
Criterio de Schwarz	211.1601	Crit. de Hannan-Quinn	201.6888

Estadísticos basados en los datos originales:

Media de la vble. dep.	46.08760	D.T. de la vble. dep.	9.168398
Suma de cuad. residuos	357.1139	D.T. de la regresión	2.915941

6.2.2. Modelo final seleccionado

Finalmente, el modelo final seleccionado es el de efectos aleatorios corregida la heterocedasticidad. Podemos observar que hay más variables significativas que en los anteriores y que el R^2 es muy cercano a uno, esto significa que nos encontramos ante un modelo en el que el desempleo queda explicado en un 98% por el modelo, pero este estimador puede ser manipulable. Por lo tanto vamos a observar el \bar{R}^2 , que no es manipulable al usar las cuasivarianzas; podemos observar que también es muy cercano a uno, por lo tanto, obtenemos el mismo resultado que con el R^2 .

En el modelo final observamos que hay cuatro variables significativas, tres de ellas al 1% y la restante al 5%.

La interpretación de las variables es la siguiente:

- Por cada punto porcentual que aumenta el abandono temprano de la educación-formación en la población de 18 a 24 años, disminuye la tasa de desempleo en menores de 25 años en un 0,56%, si el resto de variables permanecen constantes.
- Por cada punto porcentual que aumenta la tasa de paro en personas mayores de 25 años por comunidad autónoma, aumenta la tasa de desempleo en un 1,26%, si el resto de variables permanecen constantes.
- Por cada punto porcentual que aumenta la tasa de crecimiento anual del PIB a precios de mercado, aumenta la tasa de desempleo en menores de 25 años un 0,16%, si el resto de variables permanecen constantes.
- Por cada punto porcentual que aumenta la tasa de paro para la población con estudios universitarios, disminuye la tasa de desempleo en menores de 25 años en un 0,007%, si el resto de variables permanecen constantes.

- Por cada punto porcentual que aumenta la tasa de paro para la población con estudios de posgrado, aumenta la tasa de desempleo en menores de 25 años un 0,14%, si el resto de variables permanecen constantes.
- Por cada punto porcentual que aumenta la tasa de paro de larga duración para la población con estudios universitarios, disminuye la tasa de desempleo en menores de 25 años un 0,14%, si el resto de variables permanecen constantes.
- Por cada punto porcentual que aumenta la tasa de paro de larga duración para la población con estudios de posgrado, disminuye un 0,12 la tasa de desempleo en menores de 25 años, si el resto de variables permanecen constantes.

Podemos observar que algunos signos no han sido como esperábamos. La tasa de abandono esperábamos que tuviese un signo positivo, es decir, cuanto más abandono más desempleo, pero en el modelo nos sale con signo negativo. Esto puede explicarse debido a otros factores como puede ser la reinserción escolar tras el abandono, ejemplo de ello es la introducción de las Escuelas de Segunda Oportunidad (E2O) (Arredondo & Vizcaíno, 2020); otro de los motivos que puede explicar ese resultado es que el abandono temprano de los estudios a la vez supone una inserción al mercado laboral, tal y como se puede inferir de los datos, además de otros factores posibles futuras áreas de estudio. La tasa de paro de la población con formación universitaria también influye de manera negativa, por lo tanto, si aumenta el desempleo de los titulados universitarios disminuye el desempleo en los jóvenes menores de 25 años, esto puede deberse a que aunque los universitarios se encuentren más desempleados, sean los de otro tipo de formación los que si hacen que disminuya la tasa de desempleo de los menores de 25 años, como pueden ser los de la Formación Profesional Superior. Y lo mismo ocurre con la tasa de desempleo de larga duración de las personas con un grado universitario.

También hemos podido observar otros signos que si esperábamos como son los del crecimiento del PIB, la tasa de tasa de paro para la población con estudios de posgrado y la tasa de paro para los mayores de 25 años. Cuando aumenta esta última, lo hace de forma más acentuada en los menores de 25 años, como podemos observar en el gráfico 1 del Anexo. Con respecto al crecimiento del PIB, es cierto que un crecimiento de la economía suele ir acompañado de un aumento del empleo, pero no tiene por qué ser proporcional. No obstante, puede ocurrir que siga creciendo el desempleo. Por último, la tasa de paro para la población con estudios de posgrado puede hacer que el desempleo en menores de

25 años aumente porque ese grupo de población se encuentra cerca del nivel máximo de formación académica, por lo tanto, son los que a priori más oportunidades de empleo tienen.

7. Conclusiones

A través de este trabajo hemos destacado la relación que hay entre el desempleo en personas menores de 25 años y las variables escogidas, siendo las más significativas la relación negativa del abandono temprano de la formación, la positiva con la tasa de paro en mayores de 25 años y la negativa de la tasa de desempleo de larga duración para la población con estudios de posgrado. También hemos podido observar las diferencias que hay entre las comunidades autónomas y hemos descubierto un amplio campo de estudio, del cual se puede investigar e ir cambiando y reformulando el modelo, ya que es un tema actual y muy importante para el gobierno.

Para tratar de solucionar el desempleo juvenil, se han implantado medidas como son las bonificaciones por contrato, también hay de formación y para personas sin ningún tipo de formación previa. En la Unión Europea se puso en marcha la Garantía Juvenil, que es un programa que trata de facilitar el acceso a las personas entre 16 y 30 años al mercado laboral. También se ha puesto en marcha el “Plan de Garantía Juvenil Plus 2021-2027 de trabajo digno para personas jóvenes”, que va a tratar de facilitar la formación y las técnicas necesarias para entrar al mercado laboral. Esto puede contribuir positivamente a la mejora de nuestras variables explicativas, en particular al abandono temprano, y la de nuestra variable dependiente, el desempleo en menores de 25 años.

Al estar el trabajo dividido en comunidades autónomas, también puede proporcionar información a cualquier habitante de España que esté interesado en saber las posibilidades de empleo en cualquier comunidad.

8. Bibliografía

- Chica Olmo, J., & Salmerón Gómez, R. (2019). *Econometría II*. Editorial Técnica AVICAM.
- García López, J. R. (2014). El desempleo juvenil en España. *ICE, Revista De Economía*, 1(881). Recuperado a partir de <http://www.revistasice.com/index.php/ICE/article/view/1728>
- EAE Business School. (2014). *La empleabilidad de la población cualificada* (Documento 02/2014).
https://mcyt.educa.madrid.org/empleo/documentos/doc/EAE_Business_School%20La_empleabilidad_de_la_poblacion_cualificada-2013.pdf
- Perez Suarez, Rigoberto & Lopez-Menendez, Ana. (2019). *Econometria Aplicada con Gretl*.
- Montero. R (2005): *Test de Hausman*. Documentos de Trabajo en Economía Aplicada. Universidad de Granada. España
- Arredondo Quijada, R., & Vizcaíno Martínez, D. (2020). Fracaso escolar y abandono educativo temprano. Las escuelas de segunda oportunidad como alternativa. *Revista Rumbos TS. Un Espacio Crítico Para La Reflexión En Ciencias Sociales*, (23), 63-79.
<https://doi.org/10.51188/rrts.num23.423>
- Llopis Pérez, J. (2012, 13 diciembre). *Test de Belssey, Kuh y Welsch*. LA ESTADÍSTICA: UNA ORQUESTA HECHA INSTRUMENTO.
<https://jllosperez.com/2012/12/13/test-de-belssey-kuh-y-welsch/>
- Comisión Europea. (s. f.). *Iniciativa de Empleo Juvenil (IEJ)*. Empleo, Asuntos Sociales e Inclusión - Comisión Europea.
<https://ec.europa.eu/social/main.jsp?langId=es&catId=1176#:~:text=La%20Iniciativa%20de%20Empleo%20Juvenil,juvenil%20fuera%20superior%20al%2025%20%25.>

- INE. (s. f.). *Resultados por comunidades y ciudades autónomas. Serie Revisión Estadística 2019 (SEC 2010). Principales resultados.* ine.es.
https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736167628&menu=resultados&idp=1254735576581
- INE. (s. f.-a). *Abandono educativo temprano de la población de 18 a 24 años por CCAA y periodo.* INE.ES.
<https://www.ine.es/jaxi/Datos.htm?path=/t00/ICV/dim4/10/&file=41401.px>
- INE. (s. f.-c). *Tasas de paro por distintos grupos de edad, sexo y comunidad autónoma.* ine.es. <https://www.ine.es/jaxiT3/Datos.htm?t=4247>
- Salmerón Gómez, R. (s. f.). *Modelos de datos de panel Aplicaciones en ordenador.* Econometría. <https://www.ugr.es/~romansg/material/WebEco/index.html>

9. Anexo



Gráfico 1: Tasa de desempleo en los menores y mayores de 25 años

Fuente: INE

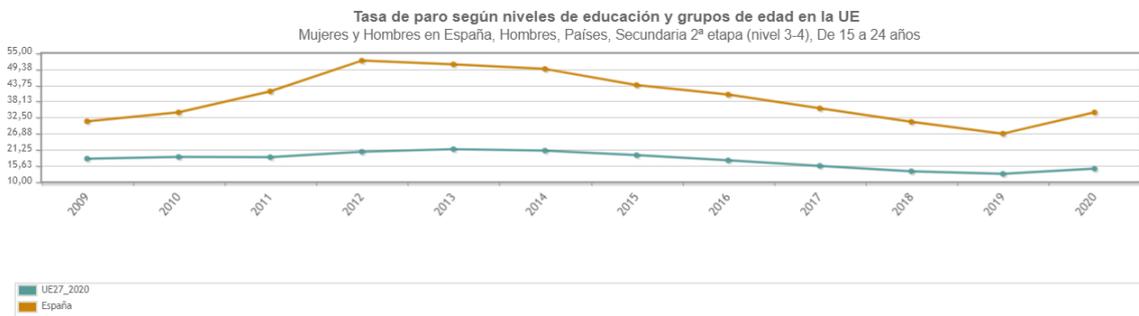


Gráfico 2: Tasa de paro de los hombres según el nivel de educación y edad entre 14 y 24 años.

Fuente: INE

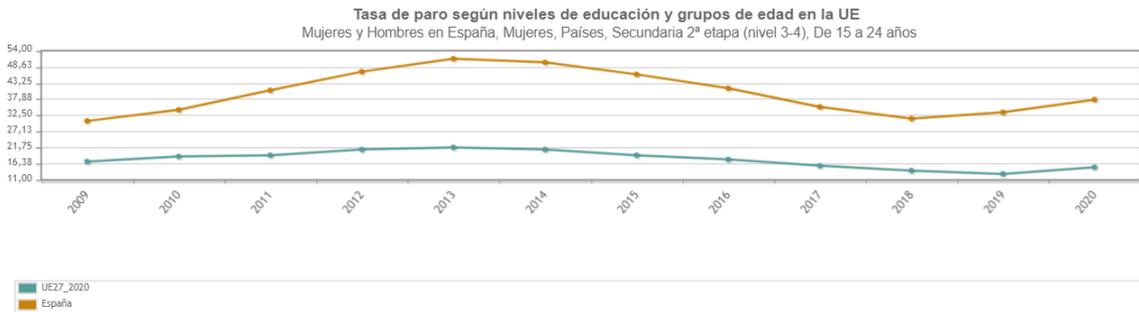


Gráfico 3: Tasa de paro de las mujeres según el nivel de educación y edad entre 14 y 24 años.

Fuente: INE