



UNIVERSIDAD
DE GRANADA

Máster Universitario en Estadística Aplicada

2021-2022

Trabajo Fin de Máster

“ANÁLISIS MULTIVARIANTE DE LOS
FACTORES QUE DETERMINAN EL
DESEMPLEO JUVENIL”

Irene García Leal

Tutor

Ramón Gutiérrez Sánchez

DECLARACIÓN DE ORIGINALIDAD

Yo, Irene García Leal con DNI 46073914S, declaro:

Que este Trabajo Fin de Máster que presento para su evaluación y defensa es original, y que todas las fuentes utilizadas para su realización han sido debidamente citadas en el mismo.

Córdoba, a 28 de junio de 2022

Firmado: Irene García Leal.

RESUMEN

Uno de los principales problemas de las economías es el alarmante nivel de desempleo que sufren los jóvenes. El objetivo del presente estudio es estudiar aquellos factores clave que afectan al desempleo juvenil en España. Para ello, se observa la evolución del desempleo juvenil en España en tiempos de crisis, observándose un aumento de paro juvenil en tiempos de recesión económica. Asimismo, se evalúa cómo ha evolucionado el desempleo juvenil español tanto a nivel nacional como entre regiones. Finalmente, se realiza un análisis multivariante, a través de un modelo de regresión lineal múltiple, de la realidad socioeconómica sumamente preocupante en las economías. La técnica de estimación empleada es un análisis de panel de datos con efectos fijos.

Palabras clave: Desempleo juvenil, España, comunidades autónomas, análisis multivariante, modelo de regresión lineal múltiple, datos de panel.

ÍNDICE GENERAL

1. INTRODUCCIÓN.	1
2. DESEMPLEO JUVENIL EN ESPAÑA: UN DIAGNÓSTICO	4
2.1. Evolución del nivel de desempleo juvenil en tiempos de crisis	4
2.1.1. El crac del 29 (1929).	4
2.1.2. La crisis del petróleo (1973)	5
2.1.3. La crisis española (1993)	5
2.1.4. La gran recesión (2008)	6
2.1.5. La crisis del COVID-19 (2020)	6
2.2. Evolución del desempleo juvenil en España: Tendencia y comparación entre regiones.	7
3. DETERMINANTES DEL PARO JUVENIL: REVISIÓN LITERARIA	11
4. ESPECIFICACIÓN ANÁLISIS MULTIVARIANTE	15
4.1. Datos y variables	15
4.2. Hipótesis	17
4.3. Técnica Multivariante	20
5. APLICACIÓN REGRESIÓN LINEAL MÚLTIPLE.	22
5.1. Análisis estadístico-descriptivo	22
5.2. Supuestos del modelo	28
5.3. Modelo Regresión Lineal Múltiple	34
6. RESULTADOS	40

7. CONCLUSIÓN Y DISCUSIÓN	43
BIBLIOGRAFÍA	47

ÍNDICE DE FIGURAS

2.1	<i>Evolución del crecimiento económico y la tasa de paro en España</i>	7
2.2	<i>Evolución de la tasa de paro juvenil en España y por Comunidad Autónoma</i>	8
2.3	<i>Evolución de la tasa de paro juvenil en las regiones del sur</i>	10
2.4	<i>Evolución de la tasa de paro juvenil en las regiones del norte</i>	10
4.1	<i>Técnicas de análisis de dependencia</i>	21
5.1	<i>Correlación entre la tasa de paro juvenil y el gasto público en educación per cápita</i>	23
5.2	<i>Correlación entre la tasa de paro juvenil y el gasto público en fomento del empleo per cápita</i>	24
5.3	<i>Correlación entre la tasa de paro juvenil y la variación poblacional menor de 25 años</i>	24
5.4	<i>Correlación entre la tasa de paro juvenil y el PIB per cápita</i>	25
5.5	<i>Correlación entre la tasa de paro juvenil y el logaritmo del PIB per cápita</i>	25
5.6	<i>Correlación entre la tasa de paro juvenil y el coste laboral</i>	26
5.7	<i>Correlación entre la tasa de paro juvenil y IPC</i>	26
5.8	<i>Correlación entre variables</i>	28
5.9	<i>Modelo de MCO agrupados (MCO)</i>	30
5.10	<i>ANOVA</i>	31
5.11	<i>Histograma de los residuos</i>	32

5.12	<i>Prueba Shapiro test</i>	32
5.13	<i>Residual vs Fitted</i>	33
5.14	<i>Plot regresión</i>	34
5.15	<i>Modelo de efectos fijos (MFE)</i>	36
5.16	<i>Modelo de efectos aleatorios (MEFA)</i>	37
5.17	<i>Pooled versus Efectos Fijos. F test for individual effects</i>	38
5.18	<i>Efectos Fijos versus Efectos Aleatorios. Test de Hausman</i>	38
6.1	<i>Modelo de mínimos cuadrados con variable dicótoma de efectos fijos</i> . .	40

ÍNDICE DE TABLAS

4.1	<i>Lista de comunidades autónomas</i>	15
4.2	<i>Descripción detallada de las variables y fuentes de datos</i>	17
4.3	<i>Tasas de paro por nivel de formación alcanzado y grupo de edad</i>	18
5.1	<i>Estadísticos resumen de las variables</i>	22
5.2	<i>Análisis de correlación</i>	27
7.1	<i>Hipótesis de partida VS Resultados del modelo</i>	44

1. INTRODUCCIÓN

El desempleo juvenil es motivo de preocupación universal. En España, el desempleo juvenil, que engloba a los menores de 25 años, es un problema estructural que se ha visto agravado en momentos de recesión, presentando graves consecuencias para la situación presente y futura de los jóvenes españoles y, en última instancia, limitando el crecimiento potencial de la economía española a largo plazo.

Este desempleo persistente se disparó considerable tras la crisis financiera de 2008. Según estadísticas oficiales de Eurostat, España alcanzó una tasa de desempleo juvenil del 56 % en 2013, y ha perdurado en el tiempo una tasa de paro juvenil del 32,5 % en 2019. La reciente crisis sanitaria, COVID-19, ha agravado considerablemente la situación laboral de la población menor de 25 años a nivel mundial y, en especial, en España. Según informa la oficina de estadística comunitaria, Eurostat, España es el país con la cifra de desempleo juvenil más elevada entre los veintisiete países de la Unión Europea (UE). De acuerdo con los datos oficiales proporcionados por este organismo, el desempleo juvenil en la UE se sitúa en el 13,9 % en marzo de 2022 respecto al 14 % del mes anterior. En términos absolutos, el número de jóvenes desempleados en la UE alcanzó en el tercer mes del año los 2.579 millones de personas, de los que 2.098 millones correspondían a la eurozona. En el caso de España, en marzo de 2022 se ha registrado 3.164 millones de personas sin empleo, de las cuales 479.000 eran menores de 25 años. De esta forma, el desempleo juvenil de España en marzo de este año ha sido de 29,6 %, la más alta de los 27 países miembros, por delante de Grecia del 27,2 % y del 24,5 % de Italia.

En el mundo laboral, ser joven se ha convertido en un gran contratiempo. Primero, por la dificultad de encontrar empleo tras llegar su fin su etapa estudiantil; y segundo, porque una vez encontrado, resulta complicado que dicho empleo presente condiciones dignas. Por otro lado, el acceso de los jóvenes al empleo se produce mayoritariamente bajo modalidades de contratación temporal. Según el Informe del Mercado de Trabajo de los Jóvenes de marzo de 2021 elaborado por el Servicio Público de Empleo Estatal (SEPE), en el año 2020, el 8,63 % del total de contratos de los jóvenes fueron de carácter indefinido

frente al 91,37 % de contratación temporal. Si comparamos en relación a 2019, estos porcentajes han variado unas décimas a favor de la estabilidad laboral de los jóvenes. En el caso de contratos de formación y contratos en prácticas, tienen una baja representación suponiendo el 0,28 % y el 0,83 %, respectivamente. Además, cabe señalar que el 53,42 % de los contratos temporales no tuvieron una duración superior a 3 meses. Finalmente, otro aspecto importante en la singularidad del modelo laboral de la juventud es la contratación con jornada a tiempo parcial. En 2018, el 43,63 % de los empleos de la juventud fueron a jornada parcial, aunque la mayor parte se trata de jornadas no deseadas pero aceptadas ante la imposibilidad de conseguir un trabajo a jornada completa (SEPE). En 2019, el tipo de jornada fue completa en el 56,20 % de los casos frente a la parcialidad del 43,80 %. Y, en 2020, la distribución ha sido del 61,58 % y 38,42 %, respectivamente. Aunque estas cifras evidencian una mejora en la calidad contractual, bien es cierto que la temporalidad y la jornada a tiempo parcial no solo implica inseguridad, sino que suele venir ligada a bajos salarios. No obstante, este trabajo no se enfoca en la precariedad laboral del empleo juvenil, siendo un debate amplio e interesante que merece un todo estudio dedicado a él.

El presente trabajo, principalmente, pretende profundizar en el estudio del desempleo juvenil en España, a través de un análisis multivariante de los factores que determinan la tasa de paro de los jóvenes menores de 25 años. En este sentido, se pretende estudiar en qué grado los diversos factores influyen en el nivel de paro juvenil para, de este modo, concienciar e informar de la precariedad existente en el mercado laboral y en qué medida afecta a la juventud.

Asimismo, el trabajo ofrece una información breve, pero a la vez suficiente, sobre la evolución del desempleo juvenil hasta la actualidad, de tal modo que sirva de orientación básica. A diferencia de la multitud de investigaciones y aportaciones en esta materia, este trabajo pretende dar no solo una visión nacional sino también de los territorios autonómicos que conforman la unidad de España. En definitiva, se busca continuar, aportar, avanzar y mejorar la obra de infinidad de economistas, sociólogos y estadistas en el desempleo juvenil.

Para alcanzar este objetivo, el trabajo se ha dividido en siete capítulos. El primero de ellos, que se corresponde con el presente epígrafe, proporciona una visión global del mercado de trabajo de los jóvenes a nivel europeo y estatal y establece los objetivos a

perseguir mediante este Trabajo Fin de Máster. El siguiente epígrafe, correspondiente al capítulo 2, ofrece un diagnóstico sobre la evolución de la tasa de desempleo juvenil en tiempos de crisis y analiza la tendencia del paro juvenil de las distintas comunidades autónomas y de España en su conjunto en los últimos años. En el capítulo 3 se muestra una síntesis de la amplia literatura existente sobre los determinantes del paro juvenil tanto a nivel nacional como internacional. En el capítulo 4 se presenta la especificación del modelo, esto es, se describen los datos y las variables empleadas, se plantean las hipótesis previas, y se decide qué técnica multivariante emplear. En el capítulo 5 se realiza un análisis estadístico-descriptivo y seguidamente, se desarrolla la metodología llevada a cabo (modelo Regresión Lineal Múltiple). A continuación, en el capítulo 6, se exponen los resultados obtenidos de la aplicación en R. Por último, en el capítulo 7, se expone la discusión de los resultados y las principales conclusiones.

2. DESEMPLEO JUVENIL EN ESPAÑA: UN DIAGNÓSTICO

2.1. Evolución del nivel de desempleo juvenil en tiempos de crisis

Los efectos de la crisis en el empleo han sido una de las principales preocupaciones de la sociedad por embestir al Estado del Bienestar. El paro cíclico es consecuencia de las crisis económicas que expulsan del mercado laboral a amplios sectores de la población, ensañándose especialmente con los jóvenes menores de 25 años y los mayores de 55 años. De acuerdo con los datos de la Encuesta de Población Activa (EPA) publicados por el Instituto Nacional de Estadística (INE), en nuestro país el paro de la población menor de 25 años y mayor de 55 años alcanzó el 54,9 % y 19,4 %, respectivamente, en el cuarto trimestre de 2013, mientras que la tasa de paro de la población comprendida entre los 25 y 54 años se situaba en torno al 24 %. Con la finalidad de soportar este argumento, en los siguientes epígrafes se aporta una breve revisión histórica sobre los efectos que las principales crisis económicas del último siglo han tenido sobre el desempleo y, en particular, sobre el desempleo juvenil, tanto a nivel nacional como internacional.

2.1.1. El crac del 29 (1929)

La bolsa de Wall Street de 1929, con un Dow Jones alcista originado por una fuerte demanda de las acciones y con altas tasas de endeudamiento para adquirirlas, unido a la falta de acompañamiento de la productividad, fue el detonante de la crisis bursátil del 29.

El impacto de esta crisis fue severa y trasladada al resto de economías del mundo. El periodo de duración fue extenso, diez años, transformándose en la Gran Depresión. En este periodo existió una elevada tasa de paro, provocada por la reducción de la producción, el hundimiento de precios y a su vez, los beneficios empresariales cayeron dando lugar a una crisis industrial y comercial.

En Estados Unidos, en 1932, la tasa de paro alcanzó el 25 % (Martín, 1981, p.7), el número de parados pasó de un millón en 1929 a trece millones en 1932. En Europa,

esta crisis no impactó con la misma intensidad; Alemania fue la más afectada llegando a alcanzar seis millones de parados en 1932, debido a que su economía dependía de los créditos exteriores.

En España, las causas de la crisis fueron varias. Una de ellas fue bursátil, pues las acciones cayeron más que la deuda pública, lo que produjo un descenso de la inversión privada y, en consecuencia, un elevado déficit comercial y una depreciación de la peseta. El paro aumentó en una España poco industrializada, focalizado en el sector agrario y de la construcción. Sin embargo, no ha sido posible determinar el alcance de esta crisis económica sobre el desempleo, ya que en este periodo no se hayan datos de cómo afectó a los desempleados menores de 25 años (los primeros datos de la EPA son del 1964).

2.1.2. La crisis del petróleo (1973)

La guerra árabe-israelí del Yom Kippur provocó un incremento del precio del crudo, debido a que los países de la OPEP no abastecieron de crudo a EE.UU. y a sus aliados durante la contienda. Este hecho, unido a la no convertibilidad del dólar en oro, fueron el origen de la crisis. Las consecuencias dieron lugar a una elevada tasa de inflación y un bajo crecimiento económico, acuñando el término de estanflación.

En EE.UU., la tasa de paro sufrió un incremento del 67 % entre 1973 y 1975, llegando a alcanzar una tasa de desempleo del 8,2 %. En España siguió un patrón similar, la tasa de paro aumentó en un 75 %, llegando a alcanzar una tasa de desempleo del 4,7 % en 1975. En aquellos años la tasa de paro era menor, comparada con la actual, fundamentalmente por dos motivos. Por un lado, la población española en edad de trabajar era inferior y, por otro lado, había más población que no buscaban trabajo, como era el caso de las mujeres. No se dispone de datos de la tasa de paro de los menores de 25 años.

2.1.3. La crisis española (1993)

La guerra del Golfo y el estallido de la burbuja inmobiliaria en Japón fueron la antesala de los acontecimientos económicos en la España de los noventa. La caída del PIB y una elevada deuda pública llevaron a la devaluación de la peseta, hasta 4 veces en nueve meses.

En 1993 el PIB disminuyó un 1 %, lo que provocó que la tasa de paro se elevase hasta el 24 %, con 3.5 millones de personas desempleadas y una tasa de paro juvenil del 43 %.

2.1.4. La gran recesión (2008)

La quiebra del banco estadounidense Lehman Brother, provocada por la concesión de hipotecas subprime o de riesgo, originó a nivel mundial una crisis financiera e inmobiliaria. El efecto de esta crisis se trasladó, también, al mercado bursátil por los productos financieros creados por los bancos de hipotecas. La fuerte concentración del dinero por grandes bancos vinculó al resto de bancos del mundo.

Las consecuencias de la crisis financiera y crisis inmobiliaria fueron la restricción del crédito, contracción de la liquidez y caída de la inversión, siendo necesaria la reestructuración y rescate de los sistemas financieros de cada país.

La crisis mundial financiera originó una contracción de la economía mundial en -0,1 % del PIB. En EEUU la tasa de paro fue del 10 %, mientras que en España alcanzó el 26 %, debido a una política monetaria tardía, auspiciada por el Banco Central Europeo (BCE). El pico más alto alcanzado por la tasa de paro juvenil fue de un 55 % en 2013.

2.1.5. La crisis del COVID-19 (2020)

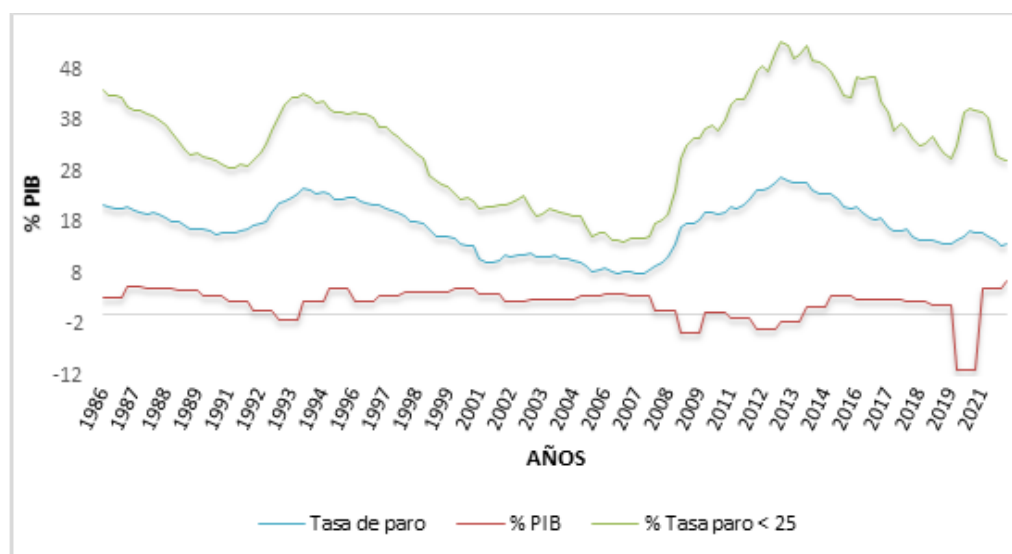
La evolución del mercado laboral de los jóvenes en 2020 se vio gravemente afectada por la crisis sanitaria Covid-19, debido al estado de alarma y el consiguiente cese de la actividad de la mayoría de sectores y la economía, en general, durante la primera mitad de 2020. Asimismo, los sucesivos rebotes en el segundo semestre de 2020 y al inicio de 2021, ha supuesto una recuperación muy leve, más de lo que se había previsto. Y es a partir del segundo semestre de 2021, gracias a la vacunación, cuando el contexto internacional y nacional era más favorable en el ámbito sanitario y, por consiguiente, en el económico. Por supuesto, que tercer y cuarto trimestre de 2021 y primero de 2022, esta tendencia de recuperación se ha visto reforzada.

Si analizamos como la crisis sanitaria a afectado al mercado laboral, se evidencia que son los jóvenes, junto con los trabajadores de baja cualificación y trabajadores temporales,

los colectivos más afectados por el impacto negativo de la crisis.

En la siguiente gráfica, se observa el denominado paro cíclico para el caso de España. Los periodos de crisis económica son seguidos de una mayor tasa de paro, existiendo una simetría inversa. La brecha más pronunciada se encuentra entre la tasa de crecimiento y la tasa de paro de los menores de 25 años durante la Gran Recesión.

Figura 2.1. *Evolución del crecimiento económico y la tasa de paro en España*

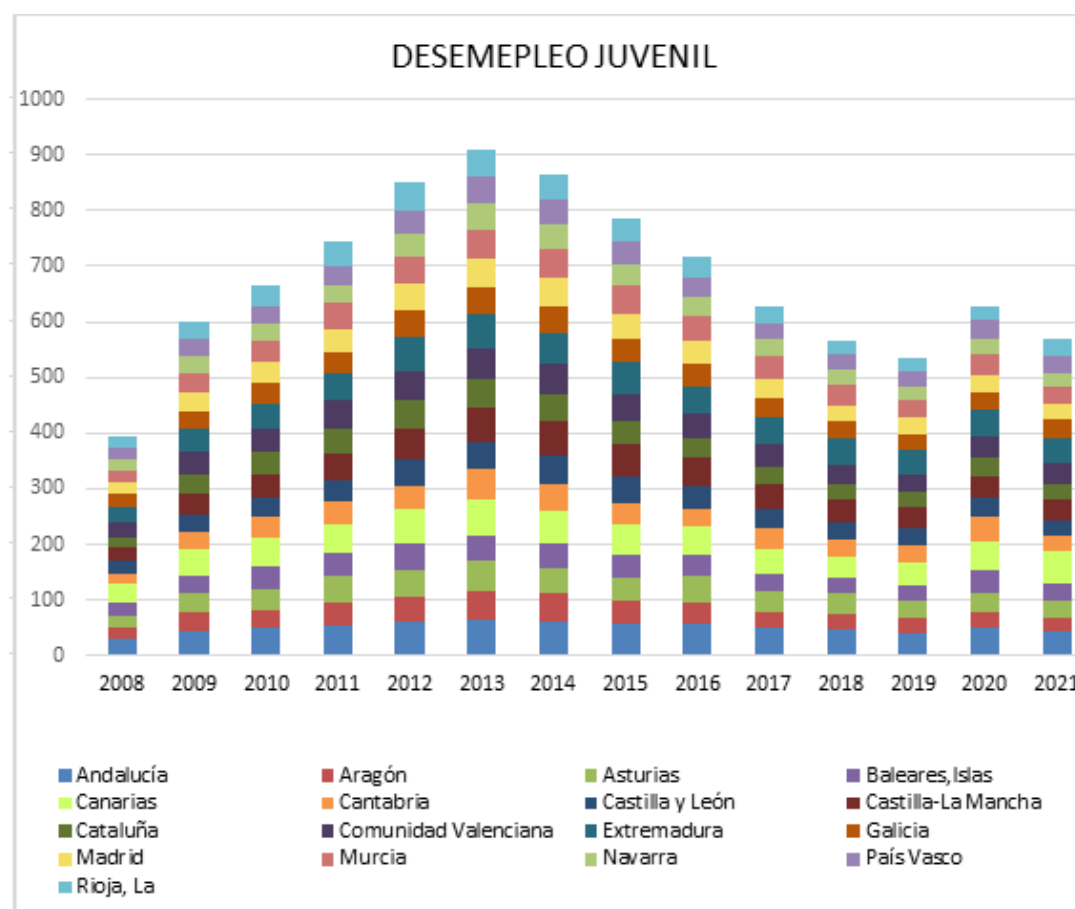


Fuente: EPA y datos macro (Expansión).

2.2. Evolución del desempleo juvenil en España: Tendencia y comparación entre regiones.

La evolución del desempleo de la población menor de 25 años hasta la actualidad a nivel autonómico sigue patrones similares a la de España en su conjunto, salvo Ceuta y Melilla. La Figura 2.2. se trata de un gráfico de barras apiladas 100 % donde se muestra el porcentaje de paro juvenil que representa cada comunidad autónoma en cada año. En línea con lo descrito en el epígrafe anterior, se observa que durante las épocas de recesión económica la tasa de desempleo alcanza sus mayores niveles. Se puede observar en la gráfica como a partir de la crisis de 2008 (Gran Recesión) la tasa de desempleo presenta una tendencia creciente hasta alcanzar el punto más álgido aproximadamente en el año 2013, tanto a nivel nacional como en cada territorio autonómico.

Figura 2.2. *Evolución de la tasa de paro juvenil en España y por Comunidad Autónoma*



Fuente: INE (2022).

Según el INE, España alcanzó su punto máximo en 2013 con una tasa de desempleo juvenil del 56,92 %. A nivel autonómico, esta tasa fue incluso superada en comunidades como Andalucía, Asturias, Islas Baleares, Canarias, Castilla-La Mancha, Extremadura y Murcia, que obtuvieron tasas de paro joven en el 2013 del 66,26 %, 63,37 %, 60,84 %, 69,23 %, 63,67 %, 65,72 % y 57,92 %, respectivamente. Sin embargo, el resto de comunidades alcanzaron sus puntos más altos en otros años, por ejemplo: Aragón alcanzó en 2014 una tasa de paro juvenil del 54,24 %; Castilla y León del 53,54 % en 2015; la Comunidad Valenciana, Madrid, Navarra y País Vasco del 62,66 %, 54,08 %, 50,41 % y 49,95 %, respectivamente, en 2014; y, por último, Ceuta alcanzó el 78,18 % en 2012 y Melilla el 75,2 % en 2015.

Si bien es cierto que la crisis financiera del 2008 deterioró a gran escala la economía española y particularmente al mercado laboral, la crisis sanitaria actual ha provocado

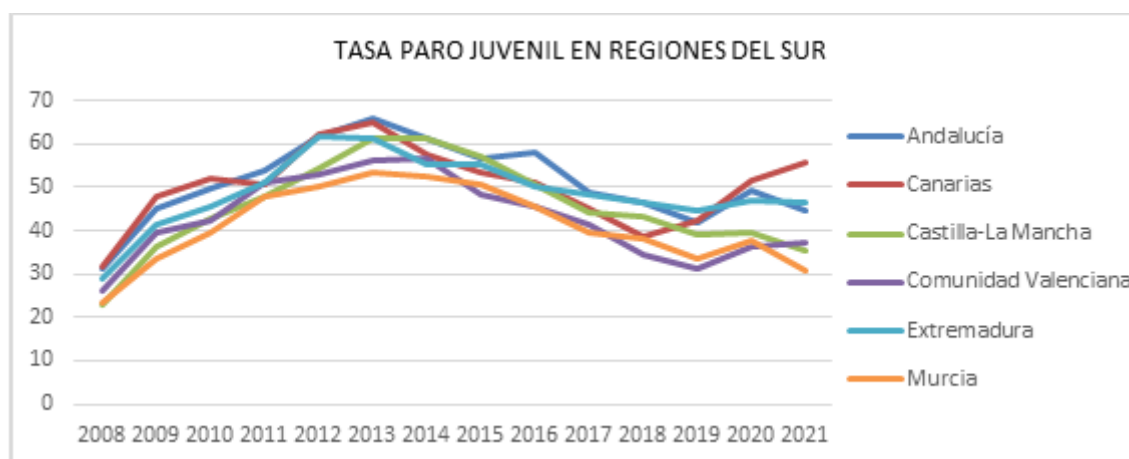
un hundimiento del empleo sin precedentes, provocando consecuencias aún más devastadoras. A diferencia de 2008, cuando la destrucción del empleo fue de manera lenta y continua durante los últimos meses, los efectos de la crisis del coronavirus se han traducido en un colapso laboral. El enorme hundimiento del empleo se produjo en menos de un mes, pues a partir del día 13 de marzo de 2020 se registró una caída en picado del paro, con niveles solo comparables con los de la Gran Recesión. Según los datos registrados en las oficinas de los Servicios Públicos de Empleo, el paro aumentó en 282.891 personas en el mes de abril de 2022 respecto al mes anterior y la Seguridad Social sufrió una pérdida de más de 500.000 ocupados en ese mes de abril, alcanzando un total de 3,8 millones de desempleados en España. Como hemos comentado, esta situación fue más grave para el caso de los parados de larga duración, los que iban a incorporarse al mercado de trabajo, los jóvenes y los de mayor edad.

Por último, señalar la existencia de fortísimas diferencias en términos de tasa de desempleo anual entre comunidades autónomas. La EPA afirma la existencia de un patrón norte-sur en cuanto al desempleo. Todas las regiones del sur de España (Castilla-La Mancha, Extremadura, Andalucía, Comunidad Valenciana, Murcia e Islas Canaria) presentaban en 2013 tasas de desempleo juvenil superior a la media nacional (que fue en torno al 55,06 %), mientras que las regiones del norte (Galicia, Asturias, Cantabria, País Vasco, Navarra, La Rioja, Aragón, Cataluña, Madrid, Baleares y Castilla y León) presentaban una tasa de paro inferior a la media. Entre los argumentos que explican dichas diferencias durante aquella recesión se encuentran: *los diferentes ciclos económicos*, pues no existe un ciclo económico sincronizado para todas las regiones, sino que existen diferentes ritmos, y con ello, procesos de destrucción de empleo distintos; *las diferentes estructuras sectoriales del empleo*, mientras que las regiones del sur de España presentan un mayor peso en la agricultura, en la construcción y en el turismo, las regiones del norte se caracterizan por la industria. Por ello, las regiones del sur fueron las más dañadas en el desempleo durante la recesión, debido a que la crisis tuvo un mayor impacto en la construcción y a la industria asociada a dicha actividad. Al igual que durante la crisis sanitaria, pues las actividades más afectadas fueron la hostelería y al turismo; *los diferentes grados de flexibilidad del mercado de trabajo*, entendido como al grado de respuesta de los salarios a las condiciones del mercado laboral; y, finalmente, *la dinámica de la población activa*, refiriéndose a la estructura de la población. Las diferencias regionales se argumentan en

que, en momentos de contracción económica, uno de los colectivos más afectados en el empleo son los jóvenes y las regiones del sur tienden a tener una población más joven que el norte de España.

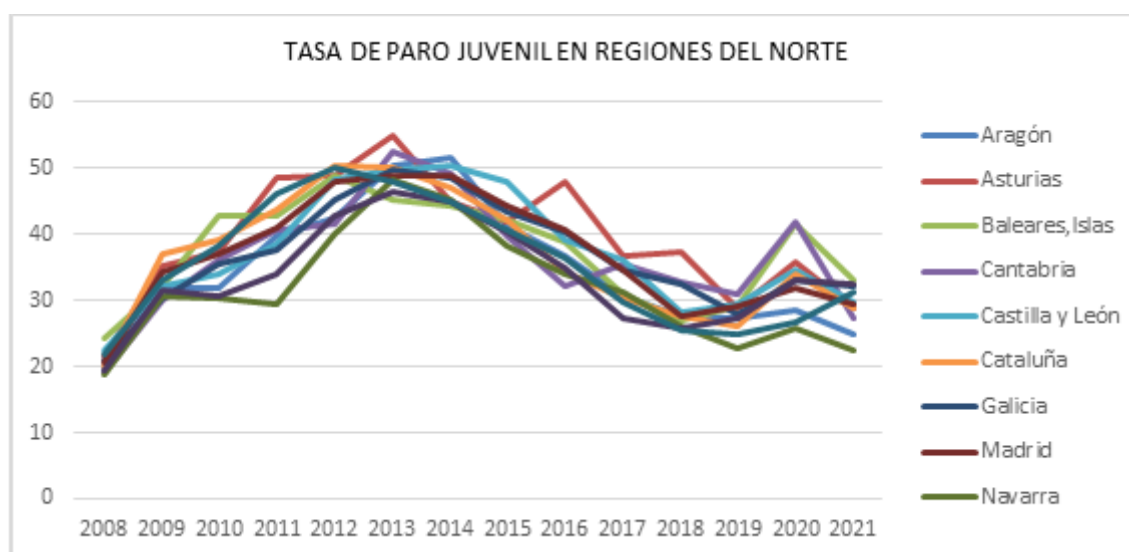
Véase en las siguientes Figuras, la evolución del paro juvenil entre el norte y sur de España, observándose como las regiones del sur presentan unas tasas de paro juvenil entre 30 y el 70 % desde 2008 a 2021 y, en cambio, la zona norte toma valores entre 20 y 50 %.

Figura 2.3. *Evolución de la tasa de paro juvenil en las regiones del sur*



Fuente: INE (2022).

Figura 2.4. *Evolución de la tasa de paro juvenil en las regiones del norte*



Fuente: INE (2022).

3. DETERMINANTES DEL PARO JUVENIL: REVISIÓN LITERARIA

La literatura sobre los jóvenes y los factores determinantes de la tasa de desempleo de este sector de la población en las últimas décadas está repleta de teorías, investigaciones, artículos, análisis cuantitativos y estudios econométricos. Algunos de los estudios más relevantes sobre los factores determinantes del desempleo, tanto para la totalidad de la población, como para el sector juvenil en particular, se señalan a continuación.

En primer lugar, comenzaremos revisando la bibliografía existente sobre el desempleo en términos generales. Por un lado, cabe señalar que el conjunto de condiciones macroeconómicas resulta ser un factor clave en la evolución de la tasa de desempleo. Así, el crecimiento del PIB es una variable explicativa fundamental para explicar los cambios en el nivel de desempleo en una determinada región. El vínculo entre ambas variables es la denominada Ley de Okun (Arthur Okun, 1962), la cual establece que, a medida que la economía crezca (la tasa de crecimiento del PIB presente valores positivos), la creación de empleo será mayor y, por consiguiente, la tasa de desempleo disminuirá. Por este motivo, se han realizado investigaciones sobre los efectos del PIB en el empleo. Por ejemplo, Lee (2000), quien confirma que la relación no es del todo estable en el tiempo y varía de un país a otro, concluyendo en todo caso que sigue teniendo validez el impacto del crecimiento en el empleo. También Solow (2000), quien utilizó una ecuación de Okun para afirmar que buena parte del desempleo en Europa se debe a la falta de demanda de trabajo. El Fondo Monetario Internacional (FMI, 2010), por su parte, estudió el papel de las instituciones y de las políticas en la explicación de los cambios en la ley de Okun en los distintos países. En 2010, la Organización Internacional del Trabajo (OIT) llevó a cabo un estudio econométrico global sobre la correlación bilateral (ILO-KILM¹, 2010) para el periodo 1992-2008, del cual se concluyó que la elasticidad del empleo global del PIB oscilaba entre 0,32 y 0,37. Bartolucci et al. (2011) detectaron, a través de un ejercicio de modelización, el impacto adicional que las crisis financieras parecen tener sobre el desempleo, debido principalmente al incremento de incertidumbre que tiene lugar en

¹ - International Labour Organization. KILM- Key Indicators of the Labour Market.

estos periodos.

En cuanto a la literatura relativa al desempleo juvenil, a continuación, se enumeran las principales conclusiones extraídas de la literatura. Al igual que ocurre con la tasa de desempleo para la totalidad de la población, se ha observado que el desarrollo económico de un territorio supone menores niveles de desempleo en los jóvenes. Así, numerosos estudios (Caporale, 2014:9; Bayrak y Tatli, 2016; Bruno et al., 2017) concluyen la evidente relación negativa que existe entre el desempleo juvenil y el PIB.

Además del PIB, existen otras variables macroeconómicas que tiene cierta influencia en el desempleo juvenil, como es la tasa de inflación. Según la conocida como curva de Phillips (1958), existe una relación inversa entre inflación y desempleo, que puede deberse a un nivel real de precios superior al esperado, lo que hace que los salarios reales sean menores a los esperados, desencadenándose un incremento del empleo y una caída del desempleo (Phillips, 1958; Solow y Samuelson, 1960). En cambio, economistas como Milton Friedman (1977) y Edmund Phelps (1967) afirmaron que la curva de Phillips solo era aplicable a corto plazo y que, a largo plazo, las políticas inflacionistas no disminuirían el desempleo. Se basaron en el fenómeno conocido como estanflación en la década de los 70, que supuso altos niveles de inflación y desempleo. Como consecuencia, las teorías basadas en la curva de Phillips fueron criticadas por muchos economistas como Thomas Sargent, Christopher Sims, Edmund Phelps, Edward Prescott, Robert A. Mundell, Robert E. Lucas, Milton Friedman y F.A. Hayek. Por otro lado, investigaciones más recientes, como el trabajo de George Akerlof, William Dickens y George Perry (2000), sugieren que existe una compensación moderada entre los bajos niveles de inflación y el desempleo, suponiendo que, si la inflación disminuye de dos a cero por ciento, el desempleo se incrementará de manera permanente en un 1,5 por ciento.

Otras variables que también influyen significativamente en las tasas de desempleo juvenil son las condiciones demográficas. Así, el porcentaje de jóvenes (o ancianos) en la población, la densidad de población, la estructura de edad de los jóvenes, los flujos migratorios (Pissarides y McMaster, 1990) y la cantidad de población y su crecimiento afectan considerablemente a las tasas de paro.

Desde el lado de la oferta de trabajo, el capital humano entre los jóvenes resulta fun-

damental en su incorporación en el mercado laboral. Así, la calidad de la mano de obra joven se considera como uno de los principales factores de desempleo juvenil (Icli, 2001; Muller, 2005) Los jóvenes con escaso capital humano y escasas cualificaciones están más expuestos al desempleo de larga duración, a la inestabilidad y a la baja calidad de los puestos de trabajo, y quizás a la exclusión social (OCDE, 2005). El problema, sin embargo, es que los jóvenes, a pesar de una educación (generalmente) superior, a menudo carecen de los otros dos componentes del capital humano: la experiencia laboral genérica y la experiencia laboral específica. De hecho, los jóvenes educados necesitan adquirir conocimientos específicos de la empresa a través de actividades laborales para que el capital humano "escolar" sea productivo (Carmeci y Mauro, 2003). Así pues, la existencia de una "brecha de experiencia juvenil" perjudica la empleabilidad de los jóvenes.

Otro aspecto interesante en la relación de los jóvenes con el empleo es el denominado fenómeno de la sobrecualificación. Dicho fenómeno afecta mayoritariamente a los jóvenes y se basa a una sobrecualificación de los mismos respecto a los trabajos que desempeñan. Según el informe elaborado por la Universidad de Valencia "El futuro de la juventud", un 22,6 % de los ocupados universitarios desempeñan un trabajo para el que están sobrecualificados, y en el caso de jóvenes con estudios superiores supone un 28 %. Una de las principales consecuencias es el desplazamiento del mercado laboral de aquellos jóvenes menos formados, agudizando el problema ocupacional de estos últimos.

De acuerdo con lo publicado en el BOE el 27/07/2013 por la Ley 11/2013, de 26 de julio, de medidas de apoyo al emprendedor y de estímulo del crecimiento y de la creación de empleo, "... existen un conjunto de debilidades estructurales que influyen directamente en las cifras de desempleo joven y sobre las que se propone trabajar, tales como la alta tasa de abandono escolar, que dobla los valores de la UE-27; la marcada polarización del mercado de trabajo, donde unos jóvenes abandonan sus estudios con escasa cualificación y otros, altamente cualificados, están subempleados; el escaso peso relativo de la formación profesional de grado medio y la baja empleabilidad de los jóvenes, especialmente en lo relativo al conocimiento de idiomas extranjeros; la alta temporalidad y contratación parcial no deseada; la dificultad de acceso al mercado laboral de los grupos en riesgo de exclusión social; y la necesidad de mejorar el nivel de autoempleo e iniciativa empresarial entre los jóvenes." Asimismo, atendiendo los objetivos propuestos a llevar a cabo por el

“Plan de Choque por el Empleo Joven 2019-2021” elaborado por Ministerio de Trabajo, Migraciones y Seguridad Social, mediante el desarrollo de las medidas oportunas: se pretende establecer un marco laboral de calidad en el empleo y la dignidad en el trabajo; se propone la prestación de una especial atención a aquellos colectivos especialmente vulnerables, tales como migrantes, parados de larga duración, abandono escolar, etc.; se persigue combatir el efecto desánimo de las personas jóvenes que no buscan empleo porque creen que no lo van a encontrar, entre otros. De este modo, resulta evidente que tanto la tasa de abandono escolar como la precariedad laboral (temporalidad y jornada parcial) son factores de gran peso en el desempleo juvenil y que se pretenden corregir a través de las diferentes estrategias de empleo joven. Es por ello que se consideran variables a tener en cuenta en el estudio.

Además de las ya comentadas, también hay otras variables políticas e institucionales que parecen ser determinantes del desempleo juvenil. Por ejemplo, Nickell et al. (2005) argumentan que el sistema de prestaciones por desempleo, el sistema de determinación de salarios, la legislación de protección del empleo, los impuestos laborales y las barreras a la movilidad laboral son variables institucionales que afectan al desempleo. En este contexto, las políticas activas del mercado laboral diseñadas de manera efectiva pueden reducir el desempleo mejorando la eficiencia del proceso de búsqueda de empleo y mejorando la experiencia laboral y las habilidades de quienes participan en ellas (Brandt et al., 2005). El papel clave que desempeñan las políticas activas del mercado laboral (ALMP) en la reducción del desempleo ha sido confirmado empíricamente en varios estudios (Scarpette, 1996; Nickell, 1997; Elmeskov et al., 1998; Choudhry et al., 2013; Bruno et al., 2017).

4. ESPECIFICACIÓN ANÁLISIS MULTIVARIANTE

4.1. Datos y variables

Con el fin de analizar el impacto de múltiples factores macroeconómicos y estructurales sobre la tasa de desempleo juvenil en España, se ha realizado un análisis multivariante de datos anuales de las 17 comunidades autónomas que conforman España para el periodo 2008-2021. Las dos ciudades autónomas (Ceuta y Melilla) quedaron excluidas del análisis por falta de datos. La lista de las comunidades autónomas incluidas en este estudio se detalla en la Tabla 4.1.

Tabla 4.1. Lista de comunidades autónomas

1. Andalucía	7. Castilla y León	13. Madrid, Comunidad de
2. Aragón	8. Castilla-La Mancha	14. Murcia, Región de
3. Asturias, Principado de	9. Cataluña	15. Navarra, Comunidad Foral de
4. Baleares, Islas	10. Comunidad Valenciana	16. País Vasco
5. Canarias	11. Extremadura	17. Rioja, La
6. Cantabria	12. Galicia	

Fuente: Elaboración propia. INE (2022).

Los datos se han obtenido principalmente de dos fuentes: el Instituto Nacional de Estadística (INE) y Expansión (datosmacro). El INE regula la actividad estadística pública, explícitamente estadísticas de gran envergadura (censos demográficos y económicos, cuentas nacionales, estadísticas demográficas y sociales, indicadores económicos y sociales...), entre otras funciones. Por otro lado, Expansión ofrece una visión global de la situación económica de cientos de países, mostrando las principales variables económicas y sociodemográficas de cada país y en cada momento. No obstante, las principales fuentes de Expansión son los organismos oficiales de los distintos países y zonas a las que corresponde los datos (INE, FMI¹, Eurostat, Banco mundial, BCE², OMC³...). Ergo,

¹FMI, Fondo Monetario Internacional.

²BCE, Banco Central Europeo.

³OMC, Organización Mundial del Comercio.

la fuente por excelencia del trabajo es el INE.

La variable dependiente es la tasa de paro juvenil (**TASA_PARO**) de 16 a 24 años, es decir, menores de 25 años. Las variables independientes son el gasto público per cápita en educación (**GP_EDUC_PC**) y en fomento del empleo (**GP_FOMENTO_PC**), la variación anual de la población menor de 25 años (**VAR_POBLA_MENOR DE 25**), el PIB per cápita (**PIBPC**), el coste laboral total por hora efectiva de trabajo (**COSTE_LABORAL**) y el índice anual de precios de consumo (**IPC**). No obstante, cabe destacar que, para conseguir un modelo válido, los datos de cada comunidad autónoma deben ser comparables entre sí. Por esta razón, la variable de gasto público en educación, gasto público en fomento del empleo, y producto interior bruto se pone en relación con la población de cada una de las comunidades autónomas (variables per cápita). En la Tabla 4.2, se muestra una explicación detallada de las variables, las unidades y la fuente de todos los datos utilizados en el análisis.

Tabla 4.2. Descripción detallada de las variables y fuentes de datos

Variable	Unidades	Descripción	Fuentes
Variable Dependiente			
TASA_PARO	Tasas (%)	Tasa De Paro Juvenil (menores De 25 Años).	EPA
Variables Independientes O Explicativas			
GP_EDUC_PC	Unidades Monetarias (€)	Gasto Público Per Cápita En Educación De Cada Comunidad Autónoma.	Expansión
GP_FOMENTO_PC	Unidades Monetarias (€)	Gasto Público Per Cápita Destinado A Políticas De Fomento Del Empleo De Cada Comunidad Autónoma.	Expansión
VAR_POBLA	Porcentaje (%)	Tasa De Variación Anual En La Población De Cada Comunidad Autónoma.	INE
PIBPC	Unidades Monetarias (€)	Producto Interior Bruto Per Cápita De Cada Comunidad Autónoma.	Contabilidad Regional De España. INE
COSTE_LABORAL	Unidades Monetarias (€)	Coste Laboral Total Por Hora De Trabajo Efectiva De Cada Comunidad Autónoma.	INE
IPC	Índice	Índice Anual De Precios De Consumo De Cada Comunidad Autónoma. Base 2021.	INE

Fuente: Elaboración propia. INE. EPA. Expansión.

4.2. Hipótesis

A continuación, se desarrolla cada una de las variables empleadas en el modelo y sus respectivas hipótesis previas.

Por un lado, la **variable dependiente** es la **tasa de paro juvenil**. Se refiere al porcentaje de jóvenes, dentro del rango de edad 16-24 años, inactivos que se encuentran en condiciones de trabajar y buscan de forma activa empleo. La tasa de desempleo de los jóvenes de un territorio está determinada por diversas variables macroeconómicas, estructurales e institucionales propias de dicha zona, es por ello por lo que cada comunidad autónoma presenta distintos niveles de desempleo.

Por otro lado, las **variables independientes** de nuestro modelo presentan, a priori, una determinada relación con la tasa de paro, permitiendo formular las siguientes hipótesis previas.

- El **gasto público per cápita en educación**, representa la partida de gasto, recogida en los presupuestos de cada comunidad autónoma, destinada a educación. Invertir en educación supone un impacto positivo en la economía de un país, ya que mejorar la calidad educativa de la población es estimular el crecimiento económico. La hipótesis previa entre ambas variables es que a más educación menor paro juvenil, es decir, una relación negativa. Bien es cierto que el paro está afectando a todos los jóvenes -incluso aquellos altamente cualificados- y el hecho de poseer un título universitario ya no es garantía absoluta de tener un trabajo. No obstante, de acuerdo con el fenómeno de la sobrecualificación, aquellos jóvenes menos cualificados son los más afectados en esta crisis ocupacional, puesto que, dada la dificultad de obtener empleo, los jóvenes mayor formación aceptan puestos de trabajos para los que están sobrecualificados, expulsando a los menos cualificados. De acuerdo con los datos del primer trimestre de 2022 del INE, la tasa de paro juvenil por nivel de estudios en España es la siguiente:

Tabla 4.3. Tasas de paro por nivel de formación alcanzado y grupo de edad

	Primera etapa de Educación Secundaria	Educación Superior
De 16 a 19 años	52,49 %	41,84 %
De 20 a 24 años	33,70 %	18,64 %

Fuente: Elaboración Propia. INE.

- El **gasto público en fomento del empleo per cápita**. Dentro de los presupuestos autonómicos existe una partida de gasto destinada a políticas de fomento del empleo. Son dotaciones para formación profesional para el empleo, oportunidades laborales, orientación profesional, fomento de la igualdad de oportunidades para el empleo, emprendimiento y mejora del marco institucional del Sistema Nacional de Empleo. Estas bonificaciones junto al *Plan de Choque por el Empleo Joven 2019-2021*, son medidas y acciones concretas que contribuyen al desarrollo y mejoran la situación laboral de los jóvenes.

La hipótesis de partida entre tasa de paro juvenil e inversión en fomento del empleo es una relación directa e inversa. Es decir, a mayor gasto público per cápita en políticas activas de empleo menor será el porcentaje de jóvenes parados.

- **Crecimiento anual de la población joven de cada comunidad autónoma.** El crecimiento de la población (porcentaje anual) representa una tasa de variación entre dos años consecutivos ($n-1$ y n) de la población de cada territorio autonómico.

La hipótesis de partida sobre el crecimiento anual de la población en relación con la tasa de desempleo juvenil es cuanto más elevada sea la variación anual poblacional mayor tasa de paro juvenil. Esta hipótesis se fundamenta en las leyes básicas de una economía de mercado (en este caso, mercado laboral): oferta y demanda, pues un crecimiento de la población supone una mayor oferta de trabajo. Dicha oferta crece a un ritmo mayor en comparación a la demanda de empleo, ergo se traduce en mayores niveles de desempleo juvenil.

- **PIB per cápita, en euros.** El producto interior bruto (PIB) per cápita es un indicador económico que mide la relación existente entre el nivel de renta de un territorio y su población. Para ello, se divide el PIB de dicho territorio entre el número de habitantes. El objetivo del PIB per cápita es indicar el nivel de riqueza o estabilidad económica de una zona en un momento determinado. Asimismo, es una medida de comparación entre diferentes países (en nuestro caso, entre las comunidades autónomas).

La hipótesis de partida es que a mayor PIB per cápita, menos paro juvenil. Es decir, cuanto mayor nivel de renta per cápita de un territorio, mayor riqueza del territorio (ignorando la distribución de la renta) y por ello, niveles de desempleo juvenil inferiores. Existe una relación inversa.

- **Coste laboral total por hora de trabajo efectiva** entendido como el coste total por hora efectiva en el que incurre la empresa por el empleo del factor trabajo. Incluye tanto el salario como las cotizaciones a la seguridad social, las percepciones no salariales (pagos por incapacidad temporal, desempleo, indemnizaciones por despido, adquisición de prendas de trabajo...), las subvenciones y bonificaciones.

La hipótesis de partida que se plantea es que la relación entre coste laboral y desempleo juvenil es positiva. A mayor coste laboral, menos contratación por parte de las

empresas y como consecuencia, mayores tasas de paro juvenil. No obstante, existen multitud de visiones y economistas que no consideran que exista tal correlación.

- **Índice de precios de consumo (IPC)** es un indicador que mide la variación de los precios de una cesta de bienes y servicios en un lugar concreto durante un periodo de tiempo determinado. Es decir, mide el impacto de las variaciones de los precios en el aumento de coste de vida. Cuando el índice es positivo significa que el IPC ha crecido y, por lo tanto, el coste de vida se incrementa, ya que los precios de productos de consumo básico han sufrido un aumento. Y, al contrario, cuando el índice es negativo significa una caída del IPC y con ello, una disminución de coste de vida puesto que los precios se han reducido. Habitualmente, se identifica IPC con inflación a pesar de no ser realmente lo mismo. Mientras que la inflación es la subida generalizada de los precios entre dos periodos concretos, el IPC es el indicador más empleado para medirla.

La hipótesis de partida entre la tasa de paro juvenil y el índice de precios al consumo es una relación negativa. La hipótesis planteada se basa en la teoría de Phillip, Solow y Samuelson sobre inflación y desempleo, la cual, tal y como ya se ha comentado anteriormente en el documento, indica una relación inversa entre ambas a corto plazo. Sin embargo, puede que la correlación entre ambas variables sea positiva, argumentándose en la teoría de Milton Friedman y Edmund Phelps, quienes afirmaron que la curva de Phillips solo funcionaba a corto plazo y que, a largo, plazo las políticas inflacionistas no reducirían los niveles de desempleo.

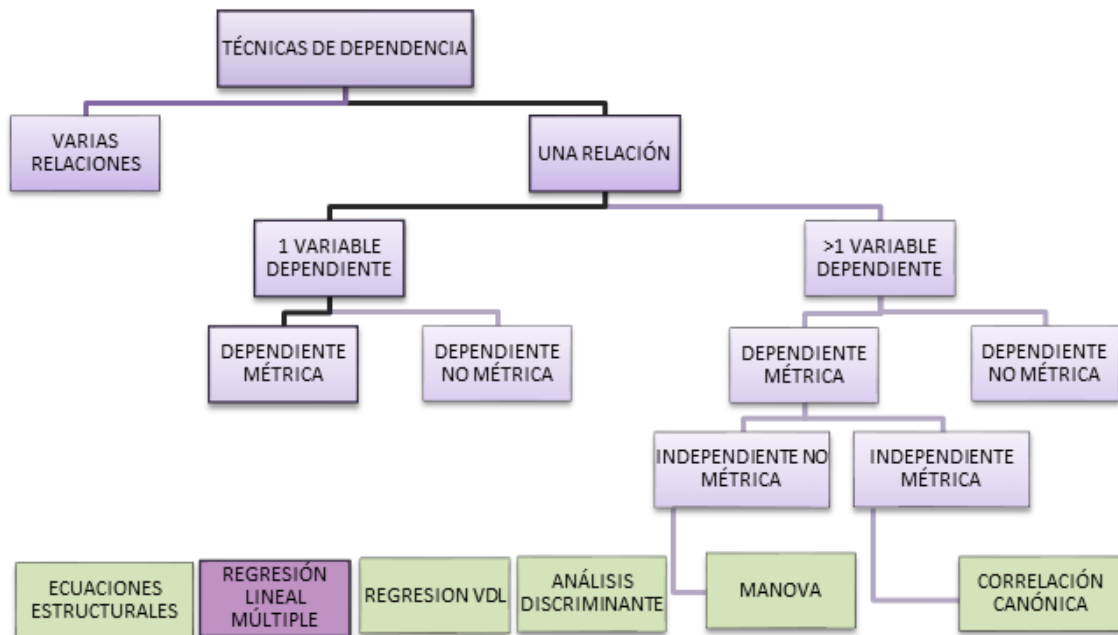
4.3. Técnica Multivariante

Una vez analizado conceptualmente el objeto de estudio, tasa de paro juvenil; definido los variables, así como identificado las relaciones fundamentales que se pretenden investigar; nos centraremos en la técnica multivariante a emplear en nuestro estudio.

Nuestro análisis se enfrenta a un problema de dependencia, cuyo objetivo es estudiar la existencia o ausencia de relaciones entre una variable dependiente – tasa de paro juvenil- y un conjunto de variables independientes. Se pretende investigar si el conjunto de determinantes afecta y en qué medida a la variable explicada de manera individual.

En la siguiente figura, se ilustra el proceso de elección de la técnica de dependencia aplicada y que atiende si se trata de un problema de dependencia o interdependencia de variables; cuántas relaciones se plantea; cuántas variables dependientes presenta nuestro estudio; y a la medida de las variables implicadas, si son métricas o no.

Figura 4.1. *Técnicas de análisis de dependencia*



Fuente: Análisis Multivariante aplicado con R, Joaquín Aldas y Ezequiel Uriel.

Elaboración propia.

Siguiendo el esquema, se aplicará un **análisis de regresión múltiple** que estudiará la relación existente entre la variable dependiente métrica, -Tasa de desempleo juvenil-, y el conjunto de variables independientes, también métricas, -Gasto público en educación, Gasto público en fomento de empleo, PIB per cápita, Variación anual de la población menor de 25, coste laboral y el IPC-.

5. APLICACIÓN REGRESIÓN LINEAL MÚLTIPLE

5.1. Análisis estadístico-descriptivo

En primer lugar, se muestra un resumen de los estadísticos descriptivos básicos de las variables originales del modelo inicial para el periodo 2008-2021. Véase en la Tabla 5.1.

Tabla 5.1. Estadísticos resumen de las variables

Variables Endógenas	N	Media	D.Típica	Mín.	Máx.	Asimetría	Curtosis
TASA_PARO (%)	238	39,73	10,44	18,80	65,99	0,17	2,36
Variables Exógenas	N	Media	D.Típica	Mín.	Máx.	Asimetría	Curtosis
GP_EDUC_PC (€)	238	858,3	145,96	648	1390	1,21	4,47
GP_FOMENTO_PC (€)	238	118,4	57,61	35	318	1,37	4,69
VAR_POBLA_MENOR DE 25 (%)	238	-0,65	2	-4,32	4,21	0,43	2,31
PIBPC (€)	238	23262	4649,27	15485	36049	0,53	2,36
COSTE_LABORAL (€)	238	19,17	2,33	14,77	27,32	0,98	3,61
IPC (Índice)	238	93,25	3,99	84,15	100,00	-0,42	2,39

Fuente: Resultados de R. Elaboración propia.

La tabla muestra datos sobre la media, la desviación típica, los mínimos y los máximos de las variables, coeficiente de asimetría y curtosis. En datos de panel, el mínimo se refiere a un individuo y un año, y el máximo a otro individuo y otro año, causando una gran dispersión entre los valores y como consecuencia, desviaciones típicas grandes. Además, debido a la amplia diferencia entre mínimos y máximos en variables como el PIB per cápita, se intuye la necesidad de aplicar logaritmos a dicha variable para facilitar el análisis del modelo. En cuanto a la variable dependiente, se aprecia la heterogeneidad de la muestra, pues presenta un valor mínimo de paro juvenil del 18,80 % y un valor máximo del 65,99 %.

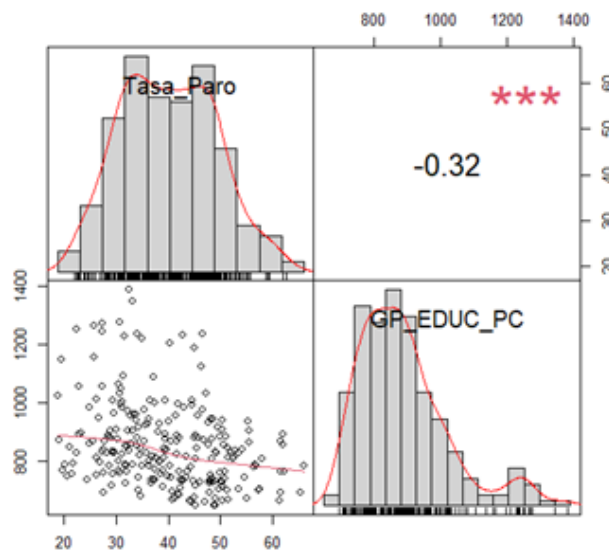
Por un lado, el coeficiente de asimetría correspondiente al gasto público per cápita en educación y al gasto público per cápita en fomento (1,21 y 1,37, respectivamente) indica una distribución sesgada positiva. Los datos sobre la variación de población joven, el PIB per cápita, el coste laboral y el IPC están moderadamente sesgados; teniendo sesgo

positivo, la variación poblacional, el PIBpc y el coste laboral (0,43, 0,53 y 0,98, respectivamente) y sesgo negativo los datos de IPC (-0,42). En cambio, los datos de la tasa de paro juvenil, la variación de población joven y el IPC son bastante simétricos, ya que tienen un coeficiente de asimetría entre -0,5 y 0,5.

En cuanto a la curtosis, medida estadística que determina el grado de concentración que presentan los valores de una variable alrededor de la zona central de la distribución de frecuencias, nos indica que todas las variables presentan una distribución leptocúrtica, es decir, más puntiaguda que la distribución normal. Esto significa que existe una más concentración de los datos en torno a la media.

Igualmente, se puede estudiar lo comentado anteriormente a través del paquete “PerformanceAnalytics”. Podemos visualizar la distribución de las variables, la nube de puntos, así como el índice de correlación. Dichos gráficos han sido extraídos de R.

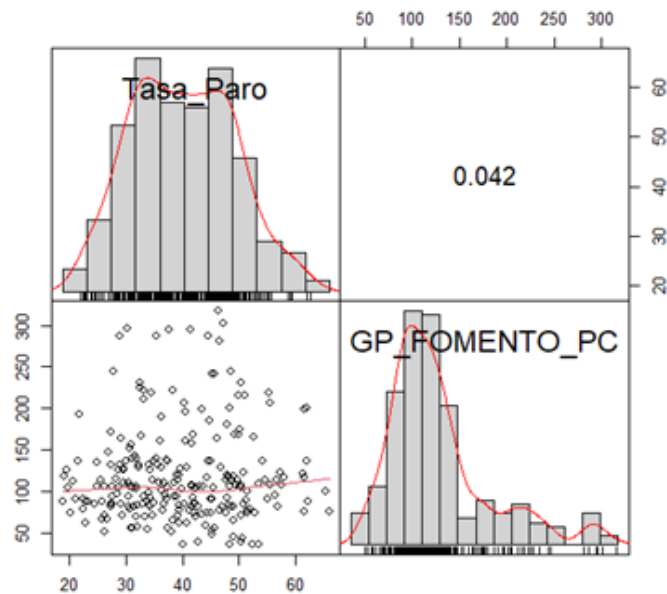
Figura 5.1. *Correlación entre la tasa de paro juvenil y el gasto público en educación per cápita*



Fuente: Resultados de R. Elaboración propia.

Como hemos comentado, la variable dependiente, -Tasa de paro juvenil-, presenta cierta simetría. Por otro lado, podemos ver como efectivamente la variable Gasto público en educación per cápita tiene una distribución sesgada positiva. El gráfico de nube de puntos nos muestra una dispersión elevada.

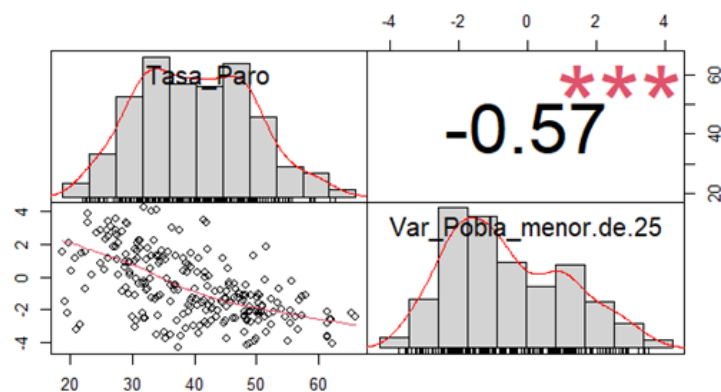
Figura 5.2. *Correlación entre la tasa de paro juvenil y el gasto público en fomento del empleo per cápita*



Fuente: Resultados de R. Elaboración propia.

En cuanto a la variable GP_FOMENTP_pc, también se evidencia una distribución sesgada hacia la derecha. Igualmente, presenta una gran dispersión entre ambas variables.

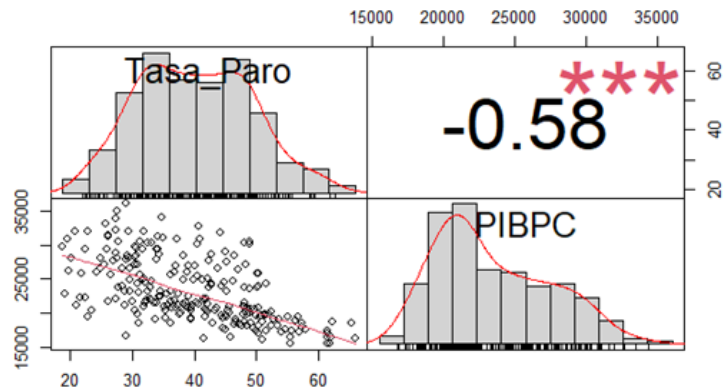
Figura 5.3. *Correlación entre la tasa de paro juvenil y la variación poblacional menor de 25 años*



Fuente: Resultados de R. Elaboración propia.

Si atendemos a la variable poblacional, observamos que presenta cierta simetría, una menor dispersión entre los datos y una distribución moderadamente sesgada positiva.

Figura 5.4. *Correlación entre la tasa de paro juvenil y el PIB per cápita*

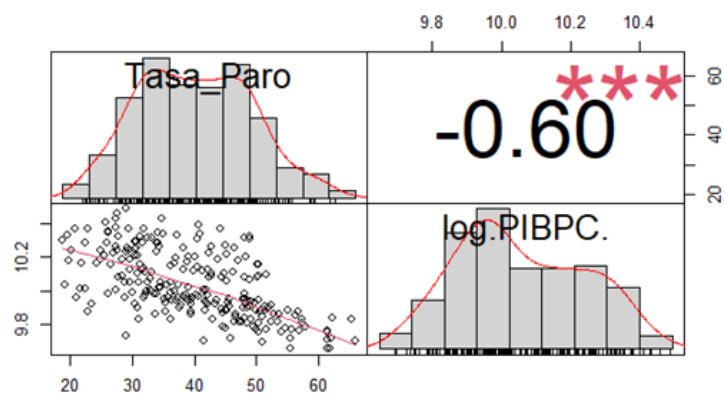


Fuente: Resultados de R. Elaboración propia.

Al igual que la variable anterior, PIBpc presenta un sesgo leve hacia la derecha y no muestra una dispersión elevada.

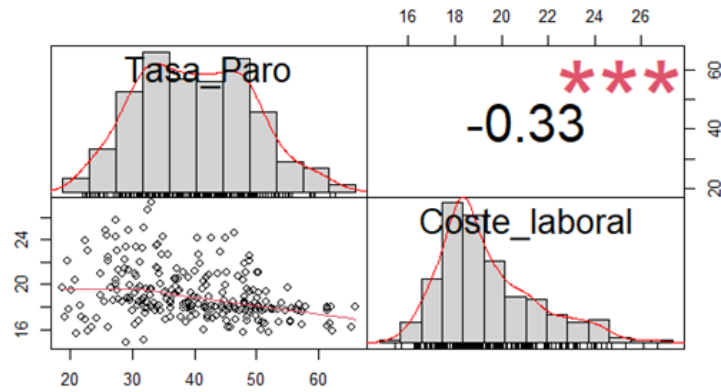
A continuación, podemos ver como la variable PIBpc una vez aplicado el logaritmo.

Figura 5.5. *Correlación entre la tasa de paro juvenil y el logaritmo del PIB per cápita*



Fuente: Resultados de R. Elaboración propia.

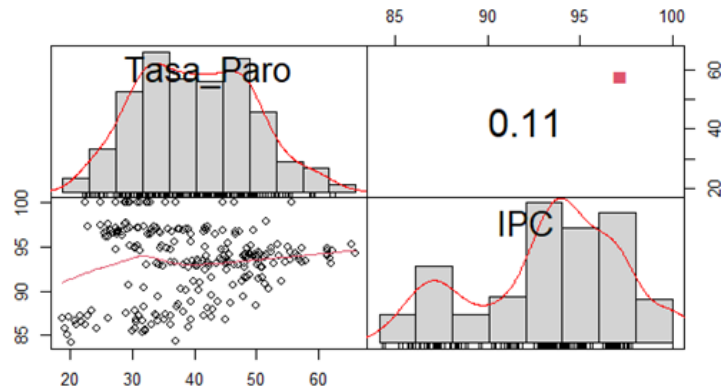
Figura 5.6. *Correlación entre la tasa de paro juvenil y el coste laboral*



Fuente: Resultados de R. Elaboración propia.

Respecto al Coste Laboral, la distribución de sus datos esta levemente sesgada hacia la derecha, al igual que la variación poblacional y el PIBpc, y la dispersión es bastante amplia.

Figura 5.7. *Correlación entre la tasa de paro juvenil y IPC*



Fuente: Resultados de R. Elaboración propia.

Finalmente, la variable IPC presenta, junto a la variable dependiente y la variación poblacional, una simetría en su compartimento. Y, a diferencia de todas, presenta un sesgo negativo.

Resumiendo, podemos decir que gracias a la nube de puntos podemos observar la correlación existente entre las variables exógenas con la tasa de paro juvenil. En el caso del PIBpc, se ha aplicado la transformación logarítmica, observándose una menor dispersión de los valores con respecto a tasa de desempleo. No obstante, los gráficos evidencian la presencia de cierta dispersión en todas las variables, pero sobre todo en el caso del IPC o el gasto público en fomento de empleo. Esto se traduce en una posible heteroscedasticidad en el modelo.

Por otro lado, la Tabla 5.2. ilustra los resultados del análisis de correlación, el cual nos permite determinar si las variables del modelo están relacionadas o no, y qué tipo de relación existe.

Tabla 5.2. Análisis de correlación

	Correlación	Probabilidad
	TASA_PARO	
GP_EDUC_PC	-0,32	0,0000
GP_FOMENTO_PC	0,04	0,5143
VAR_POBLA_MENOR DE 25	-0,57	0,0000
PIBPC	-0,58	0,0000
COSTE LABORAL	-0,33	0,0000
IPC	0,11	0,0966

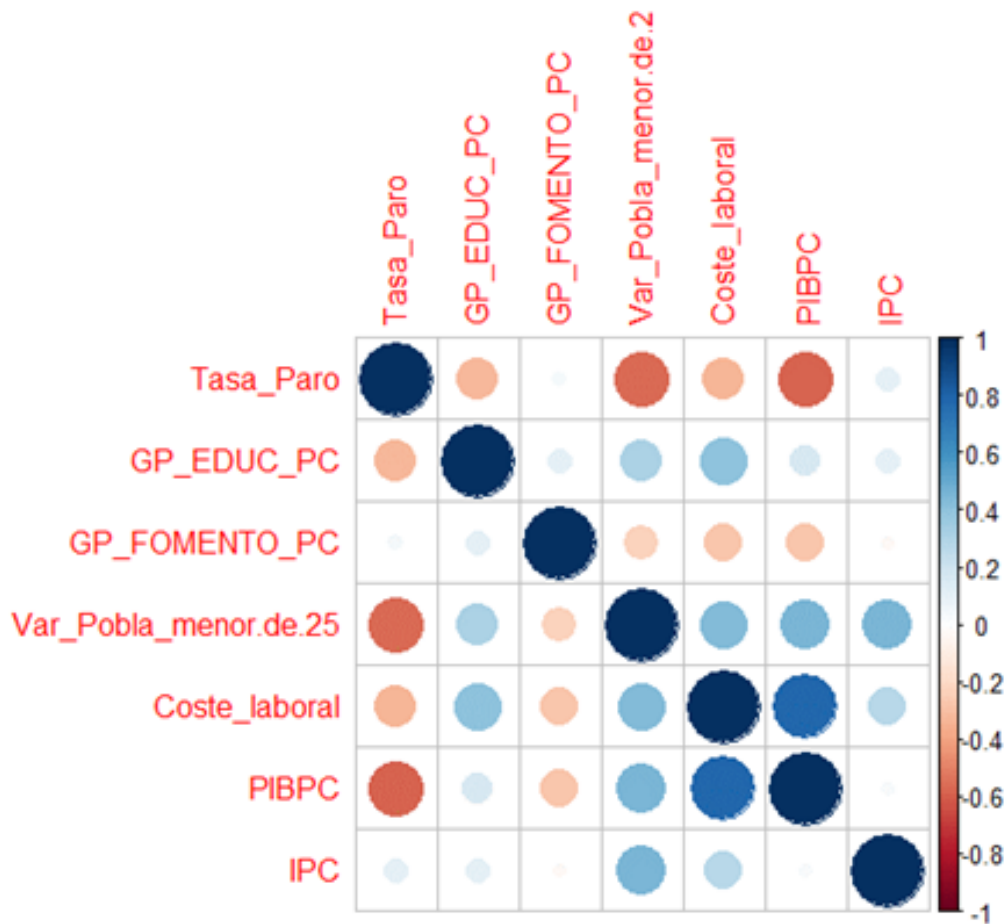
Fuente: Resultados de R. Elaboración propia.

Los valores de la segunda columna indican el coeficiente de correlación de cada variable explicativa con respecto a la tasa de paro juvenil de forma individual. Todos los coeficientes son distintos de cero y con una probabilidad, en la mayoría, menor al 5 %, lo cual significa que las variables exógenas y la endógena son dependientes significativamente. Es más, el signo del coeficiente de las siguientes variables, - gasto público en educación y PIB per cápita-, coincide con la relación planteadas en las hipótesis, epígrafe 4.2. Sin embargo, la variable sobre la inflación (IPC), la variación poblacional y el coste laboral presentan, a pesar de estar significativamente correlacionada con la tasa de paro juvenil, una correlación contraria a lo planteado en las hipótesis. Por otro lado, la variable gasto público en fomento del empleo no es significativa y muestra una relación contraria

a la planteada previamente. No obstante, más adelante, en la estimación del modelo en su conjunto, trataremos los resultados de los coeficientes de estas variables.

Gráficamente, visualizamos en la correlación de la variable dependiente con cada variable explicativa pero también la correlación existente entre variables exógenas.

Figura 5.8. *Correlación entre variables*



Fuente: Resultados de R. Elaboración propia.

5.2. Supuestos del modelo

En el modelo estimado, la tasa de paro juvenil se considera una función de todas las variables independientes desarrolladas en el epígrafe 4.1. Sin embargo, antes de decidir qué variables explicativas son las más idóneas para el modelo, se ha realizado previamente

un ajuste, reespecificando el modelo, es decir, incorporando y/o eliminando variables. En la elección de las variables, se ha tenido en cuenta que el modelo estimado tuviese una bondad de ajuste lo más óptima posible, así como que las variables fuesen significativas y, por supuesto, tuvieran una relación lo más lógica posible a las hipótesis previamente planteadas. De las veintiocho variables explicativas que inicialmente tenía la base datos, tan solo seis son las que más se ajustan al modelo. Todos los ajustes de modelo, así como los script de R, se adjuntan en el **Anexo 1**.

Por tanto, con el objeto de estudiar y analizar la importancia de cada una de las variables explicativas en la tasa de desempleo de los jóvenes se llevará a cabo un modelo de regresión lineal múltiple planteado inicialmente con las variables originales, excepto la variable PIB per cápita a la cual se le aplica logaritmos para así reducir su dispersión y mejorar la estimación del modelo.

Ecuación a estimar:

$$TASA_PARO_{it} = \beta_0 + \beta_1 GP_EDUC_PC_{it} + \beta_3 GP_FOMENTO_PC_{it} + \beta_4 VAR_POBLAMENOR25_{it} + \beta_5 LnPIBPC_{it} + \beta_6 COSTE_LABORAL_{it} + \beta_7 IPC_{it} + e_{it}$$

Con $i=1,2,\dots,17$. Se refiere a las 17 comunidades autónomas objeto de estudio. Y $t=1,2,\dots,14$. Se refiere a la serie temporal: 2008- 2021.

Donde:

- $\beta_0 \dots \beta_7$ son los coeficientes a estimar en el modelo. Los coeficientes serán la relación marginal de la variable endógena en relación a cada una de las variables exógenas y las demás ceteris paribus. Salvo el coeficiente del PIBpc que indica la elasticidad.
- e_{it} el término de error.

Es importante tener en cuenta que a la hora de trabajar con datos de panel existen varias alternativas de modelo, entre los que se encuentra: el modelo de MCO pooled o de coeficientes constantes, el modelo de efectos fijos (MEF) y el modelo de efectos aleatorios (MEFA). Sin embargo, en un primer momento, se ignorará la naturaleza de los

datos (modelo agrupado o pooled) y lo estimamos por Mínimos Cuadrados Ordinarios. En este caso, se considera que el término independiente es el mismo para todos los periodos y/o tiempos:

Figura 5.9. *Modelo de MCO agrupados (MCO)*

Call:

```
lm(formula = Tasa_Paro ~ GP_EDUC_PC + GP_FOMENTO_PC + 'Var_Pobla_menor de 25' +
Coste_laboral + log(PIBPC) + IPC)
```

Residuals:

```
Min      1Q  Median      3Q      Max
-20.2531 -3.3616  0.7454  4.3607  14.5978
```

Coefficients:

```
Estimate Std. Error t value Pr(>|t|)
(Intercept)          325.902231  41.923068  7.774 2.50e-13 ***
GP_EDUC_PC           -0.016999  0.003421 -4.969 1.31e-06 ***
GP_FOMENTO_PC       -0.022122  0.007546 -2.932 0.00371 **
'Var_Pobla_menor de 25' -2.625075  0.276375 -9.498 < 2e-16 ***
Coste_laboral         1.807020  0.338429  5.339 2.23e-07 ***
log(PIBPC)          -36.796150  3.843297 -9.574 < 2e-16 ***
IPC                   0.685827  0.129336  5.303 2.67e-07 ***
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.087 on 231 degrees of freedom

Multiple R-squared: 0.6689, Adjusted R-squared: 0.6603

F-statistic: 77.77 on 6 and 231 DF, p-value: < 2.2e-16

Fuente: Resultados de R. Elaboración propia.

Con una bondad de ajuste aproximadamente del **67 %**, las variables exógenas explican adecuadamente el modelo. Asimismo, con un p-value inferior al 5 % todos los coeficientes son significativamente distintos de 0. Puesto que las estimaciones de los coeficientes del gasto público en educación y fomento, la variación poblacional de la población menor de 25 años y el PIBpc son negativas, cuando estas variables aumentan la tasa de desempleo juvenil tiende a disminuir. En cambio, las estimaciones del coste laboral y el IPC son

positivas, esto quiere decir que la relación de dichas variables con la variable objeto van en la misma dirección, si aumenta una aumenta la otra. Finalmente, puesto que el p-valor asociado al contraste de significación conjunta es menor que 0,05, $-2,2 * 10^{-16}$, se dice que el modelo es válido conjuntamente.

A continuación, aplicamos ANOVA con el objetivo de estudiar si las variaciones en la tasa de desempleo juvenil tienen un efecto medible en cada una de las variables independientes.

Figura 5.10. ANOVA

Analysis of Variance Table

Response: Tasa_Paro

Df	Sum Sq	Mean Sq	F value	Pr(>F)
GP_EDUC_PC	1 2698.6	2698.6	72.8270	1.896e-15 ***
GP_FOMENTO_PC	1 156.2	156.2	4.2166	0.04116 *
'Var_Pobla_menor de 25'	1 6159.4	6159.4	166.2216	< 2.2e-16 ***
Coste_laboral	1 126.5	126.5	3.4144	0.06591 .
log(PIBPC)	1 7108.5	7108.5	191.8341	< 2.2e-16 ***
IPC	1 1041.9	1041.9	28.1185	2.666e-07 ***
Residuals	231 8559.8	37.1		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Fuente: Resultados de R. Elaboración propia.

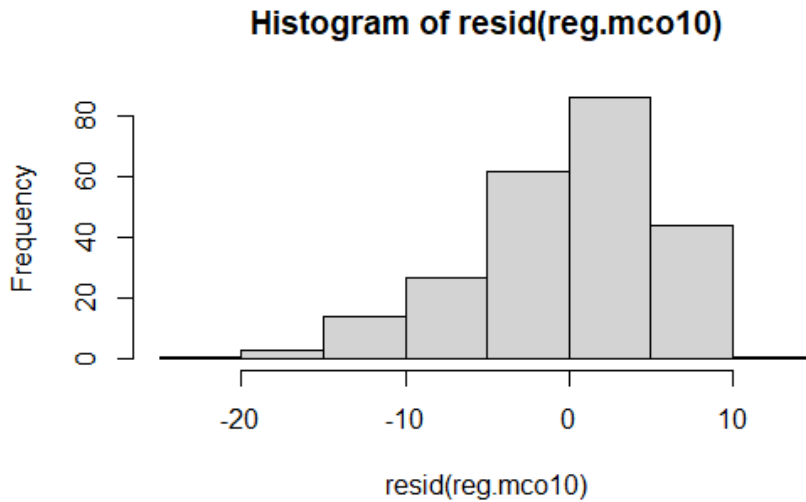
Se obtienen resultados significativos quiere decir que rechazamos la hipótesis nula y, por lo tanto, las medias son significativamente diferentes. Muy importante tener este resultado en las pruebas de la regresión, dado que esto se ajusta a lo que buscamos en la hipótesis de la regresión.

Procedemos a realizar inferencias sobre nuestras observaciones. A fin que de que estas inferencias sean “estadísticamente razonables”, han de cumplir los siguientes supuestos del modelo de regresión lineal. Resaltar que los modelos de regresión lineal son bastante “robusto”, lo que significa que no necesariamente se cumplen con exactitud las condiciones descritas a continuación.

1. **La distribución de los residuos del modelo debe ser aproximadamente normal.**

Para comprobar este supuesto, estudiaremos el histograma de residuos del modelo. Observamos que presenta una curva parecida a una distribución normal, pues

Figura 5.11. *Histograma de los residuos*



Fuente: Resultados de R. Elaboración propia.

su centro se ubica en el valor cero. Suponemos que los residuos sí muestran una distribución normal aproximadamente. Sin embargo, bien es cierto que presenta un sesgo negativo considerable. Por tanto, procedemos a evaluar la distribución de los residuos aplicando la prueba estadística de normalidad, denominada *Shapiro Wilkinson*.

Figura 5.12. *Prueba Shapiro test*

Shapiro-Wilk normality test

```
data: reg.mco10$residuals
W = 0.96937, p-value = 5.228e-05
```

Fuente: Resultados de R. Elaboración propia.

La prueba de normalidad rechaza la hipótesis nula de que los residuos se comportan de una manera normal (p-valor inferior a 0,05). Sin embargo, se ha decidido

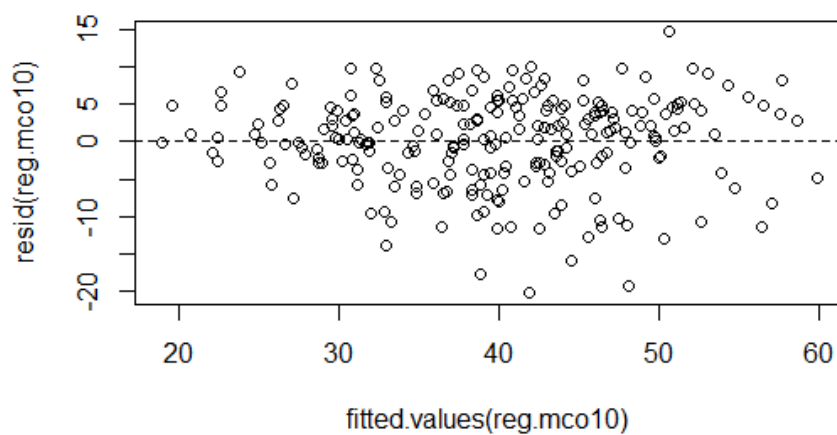
continuar con el desarrollo del nuestro estudio.

2. **La varianza de los residuos deber ser consistente para todas las observaciones. Este supuesto se conoce como Homocedasticidad.**

3. **Los residuos son independientes unos de otros.**

Estas dos condiciones se demuestran con el gráfico de valor ajustado frente a una gráfica residual.

Figura 5.13. *Residual vs Fitted*

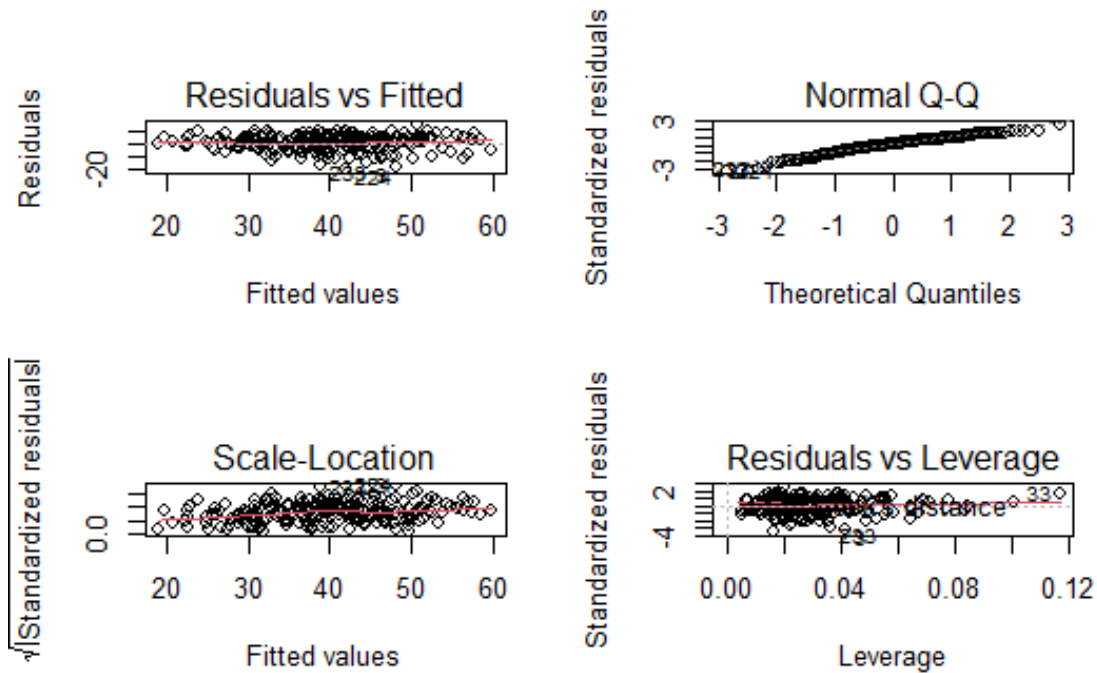


Fuente: Resultados de R. Elaboración propia.

Dado que los valores se encuentran ubicado aleatoriamente a ambos lados del valor 0, suponemos que se verifica el supuesto de homocedasticidad, es decir, tienen una varianza constante y, además, los residuos se distribuyen aleatoriamente.

Otra manera de verificar los supuestos anteriores es aplicando el comando plot, en el que muestra el gráfico de **Residual vs Fitted**; el **Normal Q-Q**, que debería sugerir que los errores residuales se distribuyen; **Scale-Location**, muestra la raíz cuadrada de los residuos estandarizados; y, por último, **Residual vs Leverage**, las distancias más grandes que 1 son sospechosos y sugieren la presencia de un valor atípico posible.

Figura 5.14. *Plot regresión*



Fuente: Resultados de R. Elaboración propia.

5.3. Modelo Regresión Lineal Múltiple

Como sabemos a la hora de trabajar con datos de panel existen varias alternativas de modelo, entre los que se encuentra: el modelo de MCO *pooled* o de coeficientes constantes, el modelo de efectos fijos (MEF) y el modelo de efectos aleatorios (MEFA). Para saber cuál es el método más adecuado para nuestro modelo, se lleva a cabo el siguiente proceso de decisión.

Inicialmente, hemos se considerado un **modelo de MCO agrupados (pooled)**, en el que se agrupan los datos de las diferentes comunidades autónomas y se estima una gran regresión, sin atender que las diferencias individuales pueden dar lugar a parámetros distintos. En este modelo inicial se asume que los parámetros son constantes y, por tanto, el término de perturbación (e_{it}) está compuesto por el error aleatorio y por un intercepto que mide el efecto individual de cada comunidad autónoma y que es constante en el tiempo.

Si se considera que el término independiente puede ser distinto para cada comunidad autónoma y/o año debe considerarse un enfoque de **efectos fijos**. Así, la estimación intra grupos (o within) puede realizarse mediante el uso de variables binarias que recojan el efecto temporal o transversal. En primer momento consideremos ambos efectos (años y comunidades autónomas).

Figura 5.15. *Modelo de efectos fijos (MFE)*

Oneway (individual) effect Within Model

Call:

```
plm(formula = Tasa_Paro ~ GP_EDUC_PC + GP_FOMENTO_PC + 'Var_Pobla_menor de 25' +
Coste_laboral + log(PIBPC) + IPC, data = mad, model = "within",
index = c("cc.AA.", "Tiempo"))
```

Unbalanced Panel: n = 19, T = 1-14, N = 238

Residuals:

Min.	1st Qu.	Median	3rd Qu.	Max.
-18.79368	-2.04529	0.38217	2.57729	11.39823

Coefficients:

Estimate	Std. Error	t-value	Pr(> t)
GP_EDUC_PC		-0.0566743	0.0056291 -10.0681 < 2.2e-16 ***
GP_FOMENTO_PC		0.0088297	0.0100830 0.8757 0.3822
'Var_Pobla_menor de 25'		0.2339197	0.2615213 0.8945 0.3721
Coste_laboral		0.8706380	0.6781026 1.2839 0.2006
log(PIBPC)		-94.7017419	6.2556507 -15.1386 < 2.2e-16 ***
IPC		0.6899764	0.1508620 4.5736 8.132e-06 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares: 17770

Residual Sum of Squares: 4765.1

R-Squared: 0.73185

Adj. R-Squared: 0.70163

F-statistic: 96.8863 on 6 and 213 DF, p-value: < 2.22e-16

Fuente: Resultados de R. Elaboración propia.

Si se considera que el término independiente es una variable aleatoria, debe considerarse un enfoque de efectos aleatorios.

Figura 5.16. *Modelo de efectos aleatorios (MEFA)*

Oneway (individual) effect Random Effect Model
(Swamy-Arora's transformation)

Call:

```
plm(formula = Tasa_Paro ~ GP_EDUC_PC + GP_FOMENTO_PC + 'Var_Pobla_menor de 25' +
Coste_laboral + log(PIBPC) + IPC, data = mad, model = "random")
```

Unbalanced Panel: n = 19, T = 1-14, N = 238

Effects:

var std.dev share

idiosyncratic 22.371 4.730 0.613

individual 14.147 3.761 0.387

theta:

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.2173 0.6814 0.6814 0.6764 0.6814 0.6814

Residuals:

Min. 1st Qu. Median Mean 3rd Qu. Max.

-15.7425 -3.0597 0.5456 0.0637 3.6342 12.4302

Coefficients:

Estimate Std. Error z-value Pr(>|z|)

(Intercept) 721.9756717 47.3516753 15.2471 < 2.2e-16 ***

GP_EDUC_PC -0.0591785 0.0048812 -12.1239 < 2.2e-16 ***

GP_FOMENTO_PC -0.0040325 0.0102658 -0.3928 0.6945

'Var_Pobla_menor de 25' -0.1659617 0.2724395 -0.6092 0.5424

Coste_laboral 3.4624004 0.4810355 7.1978 6.119e-13 ***

log(PIBPC) -70.2089537 4.8397396 -14.5068 < 2.2e-16 ***

IPC 0.0731338 0.1190862 0.6141 0.5391

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares: 19208

Residual Sum of Squares: 6725.3

R-Squared: 0.65008

Adj. R-Squared: 0.64099

Chisq: 407.885 on 6 DF, p-value: < 2.22e-16

Seguidamente, para elegir cuál de los tres enfoques anteriores es el más idóneo se disponen de los siguientes contrastes. En primer lugar, se aplica el **F-test**, con el objetivo de decidir qué modelo es más correcto, **Pooled o de Efectos Fijos**. Si se acepta la hipótesis nula el modelo agrupado es el más idóneo. En cambio, si rechazamos la hipótesis nula y, por ende, se acepta la hipótesis alternativa, preferimos el modelo de efectos fijos.

Figura 5.17. *Pooled versus Efectos Fijos. F test for individual effects*

```
F test for individual effects

data: Tasa_Paro ~ GP_EDUC_PC + GP_FOMENTO_PC + 'Var_Pobla_menor de 25' + ...
F = 9.4235, df1 = 18, df2 = 213, p-value < 2.2e-16
alternative hypothesis: significant effects
```

Fuente: Resultados de R. Elaboración propia.

Con un p-valor inferior al 0.05, se rechaza la hipótesis nula y, por consiguiente, el modelo más idóneo entre ambos es el modelo de efectos fijos.

En segundo lugar, realizamos el **Test de Hausman** para determinar si es mejor un **modelo de efectos aleatorios o un modelo de efectos fijos**, comparando las estimaciones de ambos modelos. Se pone a prueba la hipótesis nula de que las estimaciones consistentes (efectos fijos) y las eficientes (efectos aleatorios) no son significativamente diferentes. Véase los resultados en la Figura 5.18.

Figura 5.18. *Efectos Fijos versus Efectos Aleatorios. Test de Hausman*

```
Hausman Test

data: Tasa_Paro ~ GP_EDUC_PC + GP_FOMENTO_PC + 'Var_Pobla_menor de 25' + ...
chisq = 98.965, df = 6, p-value < 2.2e-16
alternative hypothesis: one model is inconsistent
```

Fuente: Resultados de R. Elaboración propia.

Con un p-valor inferior al 0.05, se vuelve a rechazar la hipótesis nula, lo que implica diferencias sistemáticas, luego es preferible el modelo de efectos fijos. Como consecuencia, se procede a estimar el modelo de datos de panel con efectos fijos, y no aleatorios. Que el modelo sea de efectos fijos quiere decir que se tiene en cuenta la heterogeneidad entre las Comunidades Autónomas, permitiendo que cada que cada unidad transversal tenga su propia variable dicótoma (intercepto). Concretamente, a este modelo se le denomina **Modelo de mínimos cuadrados con variable dicótoma de efectos fijos, MCVD** (Gujarati, y Porter, 2010).

6. RESULTADOS

Tras varios ensayos, el modelo de efectos fijos que obtiene mejores resultados considerando que el término independiente es distinto solo para cada comunidad autónoma. Luego, nuestra variable binaria recoge el efecto transversal y no temporal.

El MCVD recoge la heterogeneidad de cada sujeto en el intercepto. Por ello, utilizamos el subíndice i en el término del intercepto para indicar el valor de este para cada Comunidad Autónoma. La ecuación será la estimada en nuestro estudio, utilizando un modelo de panel de efectos fijos durante el periodo 2008-2021 para 17 Comunidades Autónomas, 238 observaciones en total.

En la siguiente salida de R se observarán los resultados del modelo estimado.

Figura 6.1. *Modelo de mínimos cuadrados con variable dicótoma de efectos fijos*

Call:

```
lm(formula = Tasa_Paro ~ GP_EDUC_PC + GP_FOMENTO_PC + 'Var_Pobla_menor de 25' +
Coste_laboral + log(PIBPC) + IPC + factor(CC.AA.))
```

Residuals:

Min	1Q	Median	3Q	Max
-14.5693	-1.8387	0.2383	2.2206	11.6538

Coefficients:

Estimate	Std. Error	t value	Pr(> t)
(Intercept)	512.192538	63.241954	8.099 4.25e-14 ***
GP_EDUC_PC	-0.048162	0.004672	-10.310 < 2e-16 ***
GP_FOMENTO_PC	-0.005814	0.008348	-0.696 0.48689
'Var_Pobla_menor de 25'	-2.405338	0.234107	-10.275 < 2e-16 ***
Coste_laboral	2.808575	0.581424	4.831 2.60e-06 ***
log(PIBPC)	-54.677262	6.403471	-8.539 2.58e-15 ***
IPC	0.664086	0.122896	5.404 1.74e-07 ***

Fuente: Resultados de R. Elaboración propia.

factor(CC.AA.)Aragón	-1.404171	3.252059	-0.432	0.66634	
factor(CC.AA.)Asturias	-15.544413	2.748436	-5.656	4.96e-08	***
factor(CC.AA.)Baleares, Islas	2.455697	2.743175	0.895	0.37169	
factor(CC.AA.)Canarias	5.155856	1.732838	2.975	0.00326	**
factor(CC.AA.)Cantabria	-0.449448	2.046770	-0.220	0.82640	
factor(CC.AA.)Castilla-La Mancha	-2.185938	1.545697	-1.414	0.15876	
factor(CC.AA.)Castilla - La Mancha	-4.818638	4.056169	-1.188	0.23617	
factor(CC.AA.)Castilla y León	-5.969766	2.304189	-2.591	0.01024	*
factor(CC.AA.)Cataluña	1.580700	4.366981	0.362	0.71774	
factor(CC.AA.)Comunidad Valenciana	3.122985	1.792978	1.742	0.08299	.
factor(CC.AA.)Comunitat Valenciana	5.100957	4.126711	1.236	0.21779	
factor(CC.AA.)Extremadura	3.225162	2.247866	1.435	0.15282	
factor(CC.AA.)Galicia	-5.145606	1.973020	-2.608	0.00975	**
factor(CC.AA.)Madrid	1.200125	5.847763	0.205	0.83759	
factor(CC.AA.)Murcia	0.642243	1.671701	0.384	0.70122	
factor(CC.AA.)Navarra	9.471407	4.419360	2.143	0.03323	*
factor(CC.AA.)País Vasco	16.848164	5.690996	2.960	0.00342	**
factor(CC.AA.)Rioja, La	2.985549	2.847457	1.048	0.29560	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.875 on 213 degrees of freedom

Multiple R-squared: 0.8763, Adjusted R-squared: 0.8624

F-statistic: 62.87 on 24 and 213 DF, p-value: < 2.2e-16

Con una probabilidad inferior al 0.05, todos los coeficientes son significativamente relevantes. Además, el modelo es significativo en su conjunto, pues el 87,63 % de las variaciones que se produce en la tasa de desempleo juvenil viene explicada por el modelo.

Los coeficientes de las distintas variables presentan resultados coherentes con las hipótesis planteadas, salvo la variable de variación poblacional menor de 25 años y de IPC. Según nuestros resultados, ante una variación de 1€ en las variables de educación

y fomento per cápita, la tasa de desempleo juvenil disminuye levemente en un 0,048 y 0,005 %, respectivamente. Respecto a la variación de la población, es una de las variables que rechaza la hipótesis de partida planteada y cuestiona la teoría de que a mayor oferta mayor paro juvenil. Pues, según el modelo, ante un crecimiento del 1 % en la población, la tasa de paro juvenil se reduciría en un 2,405 %. Por otro lado, si el PIB per cápita varía positivamente en uno 1 %, la tasa de desempleo descendería en un 54,677 %. Es decir, ante aumentos en el PIB per cápita el paro juvenil desciende debido a un mayor uso de los factores de producción, pero también, debido a que más jóvenes activos mayores niveles de renta. En cuanto a la variable del coste laboral por hora efectiva, concuerda con la hipótesis de partida y explica que, por cada variación en el coste total laboral de un 1€ el paro juvenil sufre un gran aumento del 2,809 %. Cuanto más alto el coste por trabajador, menor contratación y más aumenta el paro juvenil. Por último, si observamos a la variable del IPC, se observa una correlación positiva entre ambas variables, ante la variación de 1 punto en el IPC, la tasa de paro de los jóvenes crece un 0,664 %. La interpretación del coeficiente del IPC verifica que, en una visión a largo plazo, un incremento de la tasa de inflación no supone una reducción del nivel de desempleo en el caso de España, tal y como predijo Milton Friedman.

7. CONCLUSIÓN Y DISCUSIÓN

En esta investigación, diversos factores determinantes del desempleo juvenil en las comunidades autónomas españolas han sido estudiados a través de un análisis multivariante de datos panel. La tasa de desempleo juvenil ha sido la variable dependiente del modelo, mientras que el gasto público per cápita en educación, el gasto público per cápita en políticas de fomento del empleo, el producto interior bruto per cápita, la variación poblacional de jóvenes, el coste laboral por hora efectiva y el IPC (como medida de la tasa de inflación) han sido las variables exógenas empleadas para explicar el modelo. Se analizaron datos de las 17 comunidades autónomas españolas para el periodo 2008-2021. Se realizaron pruebas para determinar si se trataba de un modelo agrupado o anidado (F Test for individual effects), de las cuales resultó ser un modelo anidado. Seguidamente, se probaron los efectos fijos y aleatorios del modelo, concluyéndose la presencia de efectos fijos a nivel transversal.

Los resultados muestran que el gasto público per cápita tanto en educación como en fomento del empleo, el PIB per cápita y la variación poblacional afectan negativamente a la tasa de desempleo juvenil. Por el contrario, el coste laboral total y el IPC influyen positivamente en el nivel de paro juvenil.

A partir de la bibliografía de investigación, se formularon una serie de hipótesis sobre la posible influencia de las variables exógenas en la tasa de desempleo juvenil. Una vez realizado en análisis multivariante se puede observar como los resultados logrados coinciden, mayoritariamente, con de las hipótesis previamente planteadas. Para ofrecer las conclusiones y su comprensión, se facilita la siguiente tabla, que muestra un resumen de los resultados del modelo y las hipótesis de partida planteadas.

Tabla 7.1. Hipótesis de partida VS Resultados del modelo

VARIABLES INDEPENDIENTES	HIPÓTESIS DE PARTIDA	RESULTADOS DEL MODELO
Gasto público en educación per cápita	Relación negativa =	Relación negativa
Gasto público en fomento del per cápita	Relación negativa =	Relación negativa
Variación poblacional	Relación positiva ≠	Relación negativa
PIB per cápita	Relación negativa =	Relación negativa
Coste Laboral	Relación positiva =	Relación positiva
IPC	Relación negativa ≠	Relación positiva

Elaboración propia.

La educación es un factor muy importante en el mercado laboral, y así lo demuestra los resultados del modelo, donde un mayor gasto público en educación influye de forma directa y negativa en la tasa de paro juvenil. Por ello, invertir en educación es, indudablemente, una medida para combatir el desempleo juvenil. No obstante, existen múltiples factores que alientan el abandono escolar, uno de ellos son los bajos niveles salariales que sufren los jóvenes, pues tener más formación no se traduce, en muchos casos, tener mejor salario. Aún más, debido a que el sistema productivo no es capaz de integrar el talento de los jóvenes, se ven en la necesidad de emigrar a otros países en busca del empleo que no encuentran en España, una realidad que no estimula la formación profesional.

Respecto al gasto público en fomento del empleo, se han obtenido resultados acordes con la hipótesis planteada. Una mayor inversión en políticas activas para el empleo ayudará a solucionar el problema crónico del desempleo juvenil del que padece España. Desde que se agudizó el desempleo entre los jóvenes tras la crisis financiera, se han ido desarrollando estrategias para combatir las altas tasas de desempleo, para poner fin a la precariedad laboral en el empleo joven, crear oportunidades de trabajo que suponga la inserción laboral de los jóvenes, con garantías del acceso a puestos de trabajo dignos y salarios decentes. Destinar recursos económicos en planes de fomento de empleo juvenil es un aspecto que no solo beneficia al propio colectivo en cuestión, sino a la sociedad en su conjunto, ya que son el futuro del país.

En lo que a la variación poblacional se refiere nos encontramos con resultados no esperados. La hipótesis previa se regía por las leyes de oferta y demanda, donde un aumento

de la población significa un aumento de la oferta de trabajo (aumento de la población activa) y, por lo tanto, mayores niveles de desempleo juvenil. Los resultados obtenidos harían rechazar la hipótesis inicial y cuestionar la teoría de oferta y demanda del mercado laboral. No obstante, la relación negativa entre la tasa de paro juvenil y la variación poblacional resultante del modelo puede justificarse en que una variación positiva de la población viene determinada por los nacimientos y defunciones registrados, pero también por los flujos migratorios efectuados. Así, si la población de un determinado territorio aumenta puede estar vinculada a la llegada de personas en busca de trabajo. ¿Y por qué mayor población menor desempleo? Podría ser porque el hecho de que las personas que deciden movilizarse por motivos laborales lo hagan principalmente porque en esa zona hay mayor demanda de trabajo y, por lo tanto, mayores oportunidades de empleo, que conducirá a una reducción de paro.

El PIB es una variable estrechamente vinculada con la creación de empleo. Es por ello, que la variable del modelo, PIB per cápita, concuerda con la hipótesis de partida. Ante un crecimiento económico, la tasa de paro juvenil disminuiría debido a un mayor empleo de los factores de producción y a que los jóvenes generan más renta. En cambio, en cuanto al IPC, los resultados del modelo discrepan la teoría de la Phillips y podrían justificarse en la tesis de Milton Friedman sobre la pendiente positiva en la curva de Phillips a largo plazo. Los resultados muestran estar más línea con las aportaciones de economistas a finales de los 80 en épocas de recesión como ocurre actualmente. Pues si consideramos el estancamiento real de la economía, la relación positiva del IPC y la tasa de desempleo se argumenta en que un incremento generalizado de los precios no se traduce en una disminución de desempleo, ya que los salarios en la mayoría de los sectores se encuentran congelados y por ello, el resultado de esa subida de precios es una reducción del poder adquisitivo de la población, desencadenándose en una reducción del consumo y con ello, destrucción del empleo.

Por último, en el caso de la variable del coste laboral al demuestra ir en línea con las hipótesis de partida. Atendiendo a la precariedad laboral existente, los contratos laborales y a tiempo parcial son un componente clave en este tipo de “empleo basura”. Se tratan de puestos de trabajo que obstaculizan el crecimiento laboral y personal de la juventud, ya que no proporcionan inestabilidad económica, dificultando la emancipación de la po-

blación joven, entre innumerables consecuencias. Pero además de no tener empleos de calidad, los jóvenes son los más vulnerables laboralmente en momentos de recesión. Esto se debe a que el primer perfil expulsado del mercado laboral durante una crisis son los trabajadores temporales, que mayoritariamente son empleos cubiertos por la población menor de 25 años. Por este motivo, resulta fundamental aplicar medidas que fomenten el empleo de calidad y digno para los jóvenes.

Finalmente, el objetivo de este trabajo no es otro que continuar el trabajo de multitud de economistas, aportando una visión cuantitativa de los factores que influyen en los niveles de desempleo juvenil. Ahondar en la preocupante situación laboral de los jóvenes, con el fin de reducir el alto volumen de desempleo juvenil y acabar con la precariedad laboral.

BIBLIOGRAFÍA

Akerlof, George A.; Dickens, William T.; Perry, George L. (2000). “Salario casi racional y fijación de precios y la curva de Phillips a largo plazo”. Documentos de Brookings sobre actividad económica, 2000 (1), pp. 1–60. CiteSeerX 10.1.1.457.3874. doi: 10.1353 / eca.2000.0001.

Bartolucci, F., Choudhry, M., Marelli, E. and Signorelli, M. (2011). “Financial Crises and Labour Market: Beyond the Okun’s Law”, presented at the 16th World Congress International Economic Association, Tsinghua University, Beijing, July 4-8, 2011; also at the 26th National Conference of Labour Economics (AIEL), Università Cattolica del Sacro Cuore, Milan, 15-16 September 2011.

Bayrak, R., Tatli, H. (2016). “Short and long-term analysis of some factors affecting youth unemployment in Turkey”, *Theoretical and Applied Economics*, 13[3(608)], 229-242.

Bayrak, R., Tatli, H. (2018). “The Determinants of Youth Unemployment: A Panel Data Analysis of OECD Countries”, *The European Journal of Comparative Economics*, 15 (2), pp. 231-248.

Brandt N., Burniaux J. M., Duval R. (2005). “Assessing the OECD Jobs Strategy: Past Developments and Reforms”, OECD Economics Department Working Paper, 429.

Bruno, G. S. F. et al. (2017). “The Short and Long-run Impacts of Financial Crises on Youth Unemployment in OECD Countries”, *Applied Economics*, 2017, 49(34), 3372-3394.

Caporale, G. M. (2014). “Youth Unemployment in Europe: Persistence and Macroeconomic Determinants”, CESIFO Working Paper, 4696, <https://ideas.repec.org/p/ces/ceswps/_4996.html>(último acceso 27/06/2022).

Carmeci L. and Mauro L. (2003). “Long Run Growth and Investment in Education: Does Unemployment Matter?”, *Journal of Macroeconomics*, 25, 123-137.

Choudhry M. T., Marelli E. y Signorelli M. (2013). “Youth and the Total Unemployment Rate: The Impact of Policies and Institutions”, *Rivista Internazionale di Scienze*

Sociali, 121(1), 63-86).

Elmeskov J., Martin J. P. y Scarpetta S. (1998). “Key Lessons for Labour Market Reforms: Evidence from OECD Countries’ Experiences”, *Swedish Economic Policy Review*, 5(2), pp. 205–52.

Eurostat (2022): Unemployment by sex and age - annual data, <https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=une_rt_a&lang=en>(último acceso 27/06/2022)

Friedman, M. (1977). “Inflation and Unemployment”, *The Journal of Political Economy*, 85(3), 451-472.

García-Meza, M.A. (2018). “utorial de Regresion con Panel de Datos” <<https://rpubs.com/MariusAgm/Panel>>(último acceso 27/06/2020)

Gerring, J. (2006). *Case Study Research: Principles and Practices*. Cambridge: Cambridge University Press.

Gujarati, Damodar N y Porter, Dawn C. (2010). “Modelo de regresión con datos de panel”. *Econometría*, 5ª ed., Editorial MacGrawHill., pp.591-616.

IMF (2010). “Unemployment Dynamics During Recessions and Recoveries: Okun’s Law and Beyond”, Chapter 3 in *World Economic Outlook: Rebalancing Growth*, Washington.

ILO (2010). “Global Employment Trends for Youth. Special Issue on the Impact of Global Economic Crisis on Youth”, International Labour Organization, Geneva, August.

ILO (2012). “Key Indicators of the Labour Market (KILM)”, Seventh Edition, Geneva. Içli G. (2001). “Education, Employment and Technology”, *Pamukkale University Journal of Education*, 9, pp. 65-71.

Lee, J. (2000). “The Robustness of Okun’s law: Evidence from OECD Countries”, *Journal of Macroeconomics*, 22(2), pp. 331-56.

Ley 11/2013, de 26 de julio, de medidas de apoyo al emprendedor y de estímulo del crecimiento y de la creación de empleo. *Boletín Oficial del Estado*, 27 de julio de 2013.

Lieberman, E. S. (2005). Nested analysis as a mixed-method strategy for comparative

research. *American Political Science Review*, 99(3), 435-452.

Martín, I. (1981). "Crisis económicas del Siglo XX", Salvat Editores, S.A., Barcelona.

Martinez-Valdes, G. (2022). "Regresión lineal múltiple", <<https://rpubs.com/gustavomtzv/903768>>(último acceso 27/06/2020)

Muller W. (2005). 'Education and Youth Integration into European Labour Markets', *International Journal of Comparative Sociology*, 46(5), pp. 461-485.

Nickell, S. J. (1997), 'Unemployment and Labour Market Rigidities: Europe versus North America', *Journal of Economic Perspectives*, 11(3), pp. 55–74.

Nickell S., Nunziata L. y Wolfgang O. (2005), 'Unemployment in the OECD since the 1960s. What do we know?', *Economic Journal*, 115, 500: 502-06.

OECD (2005). "Education at Glance", Paris.

Okun, A. M. (1962). "Potential GNP: Its Measurement and Significance". *American Statistical Association*, pp. 89-104.

Phelps, Edmund S. (1967). "Curvas de Phillips, expectativas de inflación y desempleo óptimo en el tiempo", *Economica*, 34(135), pp. 254–281.

Phillips, A.W. (1958). "The Relation Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861-1957", *Economica*, 25(100), pp.283-299.

Ramirez-Alán, O. (2017). "Correlación y Regresión Lineal" <<https://rpubs.com/osoramirez/316691>>(último acceso 27/06/2020)

Salmerón Gómez, R. (2017). "Entorno de programación RStudio: regresión con datos de panel" <https://www.ugr.es/romansg/material/WebEco/04-Eco2/Ordenador/R/03_PanelData.html>(último acceso 27/06/2020)

Seawright, J. (2016). *Multi-Method Social Science: Combining Qualitative and Quantitative Tools*. Cambridge: Cambridge University Press.

Secretaría de Estado de Empleo y Economía Social, Subdirección General de Estadís-

tica y *Á*lisis sociolaboral (2022). “Informe Jóvenes y Mercado de Trabajo Marzo 2022”
<https://www.mites.gob.es/ficheros/ministerio/sec_trabajo/analisis_mercado_trabajo/jovenes/2022/Marzo_2022.pdf>(último acceso 27/06/2020)

SEPE (2022): Informe del Mercado de Trabajo de los Jóvenes 2022 (Datos 2021), <file:///C:/Users/USUARIO/Downloads/Informe-Mercado-Trabajo-Jovenes-2022-datos2021.pdf>(último acceso 27/06/2022).

Scarpetta S. (1996), ‘Assessing the role of labor market policies and institutional settings on unemployment: a cross-country study’, *OECD Economic Studies*, 26, 43–98.

Solow, R.M. y Samuelson, P.A. (1960). “Analytical Aspects of Anti-Inflation Policy”, *American Economic Review*, 50(2), pp. 177-194.

Solow, R.M. (2000). ‘Unemployment in the United States and in Europe: A Contrast and the Reasons’, *CESifo Working Paper*.

ANEXO 1

```
#IMPORTAR DATOS
```

```
library(tidyverse)
```

```
library(fs)
```

```
library(readxl)
```

```
setwd("F:/FORMACIÓN/Máster Estadística Aplicada/CURSOS/TFM/BASE DE DATOS")
```

```
read_excel("Database1.xlsx")
```

```
read_excel("Database1.xlsx", 2)
```

```
read_excel("Database1.xlsx", 3)
```

```
#APLICAMOS LA FUNCIÓN EXCEL_SHEETS() PARA EXTRAER LOS NOMBRES DE LAS HOJAS
```

```
path <- "Database1.xlsx"
```

```
path %>%
```

```
excel_sheets()
```

```
mad <- path %>%
```

```
excel_sheets() %>%
```

```
set_names() %>%
```

```
map(read_excel,
```

```
path = path)
```

```
str(mad)
```

```
#UNIMOS LAS TABLAS POR FILA
```

```
mad <- path %>%
```



```
excel_sheets() %>%
```

```
set_names() %>%
```

```
map_df(read_excel,
```

```
path = path)
```

```
#NUESTRA BASE DE DATOS
```

```
mad
```

```
attach(mad)
```

```
#ANÁLISIS DESCRIPTIVO
```

```
summary(mad)
```

```
sd(Tasa_Paro)
```

```
sd(GP_EDUC_PC)
```

```
sd(GP_FOMENTO_PC)
```

```
sd('Var_Pobla_menor de 25')
```

```
sd(Coste_laboral)
```

```
sd(PIBPC)
```

```
sd(IPC)
```

```
library(moments)
```

```
skewness(x) #nos da el valor de la asimetria de los datos de la variable x
```

```
kurtosis(x) #nos da el achatamiento de la distribucion de los datos  
de la variable x.
```

```
skewness(Tasa_Paro)
```

```
skewness(GP_EDUC_PC)
```

```
skewness(GP_FOMENTO_PC)
```

```
skewness('Var_Pobla_menor de 25')
```

```
skewness(Coste_laboral)
```

```
skewness(PIBPC)
```

```
skewness(IPC)
```

```
kurtosis(Tasa_Paro)
```

```
kurtosis(GP_EDUC_PC)
```

```
kurtosis (GP_FOMENTO_PC)
```

```
kurtosis('Var_Pobla_menor de 25')
```

```
kurtosis(Coste_laboral)
```

```
kurtosis(PIBPC)
```

```
kurtosis(IPC)
```

```
# cor.test analiza la significancia de la correlacion, utilizando el  
# contraste t-student.
```

```
cor.test(Tasa_Paro, GP_EDUC_PC)
```

```
cor.test(Tasa_Paro, GP_FOMENTO_PC)
```

```
cor.test(Tasa_Paro, 'Var_Pobla_menor de 25')
```

```
cor.test(Tasa_Paro, Coste_laboral)
```

```
cor.test(Tasa_Paro, PIBPC)
```

```
cor.test(Tasa_Paro, IPC)
```

```
dat1 <- data.frame(Tasa_Paro, GP_EDUC_PC)
```

```
chart.Correlation(dat1)
```

```
dat2 <- data.frame(Tasa_Paro, GP_FOMENTO_PC)
```

```
chart.Correlation(dat2)
```

```
dat3 <- data.frame(Tasa_Paro, 'Var_Pobla_menor de 25')
```

```
chart.Correlation(dat3)
```

```
dat4 <- data.frame(Tasa_Paro, Coste_laboral)
```

```
chart.Correlation(dat4)
```

```
dat5 <- data.frame(Tasa_Paro, PIBPC)
chart.Correlation(dat5)
```

```
dat7 <- data.frame(Tasa_Paro, log(PIBPC))
chart.Correlation(dat7)
```

```
dat6 <- data.frame(Tasa_Paro, IPC)
chart.Correlation(dat6)
```

```
M <- cor(datosf) #permite ejecutar una matriz de correlacion
```

```
corrplot::corrplot(M, method = "circle")
```

```
# AJUSTES REGRESIÓN MÚLTIPLE LINEAL. MODELO POOLED.
```

```
reg.mco1 = lm(Tasa_Paro~'GP_EDUC_%GP'+ 'GP_FOMENTO_%GP'+ '%
POB_<25_S/TOTAL'+ 'PIB%'+ Coste_laboral+'DEUDA_%PIB'+ 'IPC')
summary(reg.mco1)
```

```
reg.mco2 = lm(Tasa_Paro~'GP_EDUC_€'+ 'Pobla_menor de 25'+ 'PIB%'+
Coste_laboral+'DEUDA_TOTAL'+ 'IPC')
Summary(reg.mco2)
```

```
reg.mco3 = lm(Tasa_Paro~'VAR_GP_EDUC'+ 'Var_Pobla_menor de 25'+
'VAR_DEUDA'+ 'VAR_IPC'+ Coste_laboral)
summary(reg.mco3)
```

```
reg.mco4 = lm(Tasa_Paro~'VAR_GP_EDUC'+ 'Var_Pobla_menor de 25'+
```

```
'VAR_DEUDA TOTAL'+ 'IPC'+ log(Coste_laboral))
summary(reg.mco4)
```

```
reg.mco5 = lm(Tasa_Paro~'VAR_GP_EDUC'+ 'Var_Pobla_menor de 25' +
'VAR_DEUDA%'+ 'IPC'+ log(Coste_laboral))
summary(reg.mco5)
```

```
reg.mco6 = lm(Tasa_Paro~'VAR_GP_EDUC'+ 'Var_Pobla_menor de 25' +
'VAR_DEUDA%'+ VAR_IPC+ log(Coste_laboral))
summary(reg.mco6)
```

```
reg.mco7 = lm(Tasa_Paro~log('GP_EDUC_€')+ log('GP_FOMENTO_€')
+log('Pobla_menor de 25') + log(DEUDA_TOTAL)+ 'IPC'+ log(Coste_laboral))
summary(reg.mco7)
```

```
reg.mco8 = lm(log(Tasa_Paro)~log('GP_EDUC_€')+log('Pobla_menor de 25') +
log('DEUDA_%PIB')+ log('IPC')+ log(Coste_laboral))
summary(reg.mco8)
```

```
reg.mco9 = lm(log(Tasa_Paro)~log('GP_EDUC_€')+log('Pobla_menor de 25') +
log('DEUDA_%PIB')+ IPC+ log(Coste_laboral))
summary(reg.mco9)
```

#MODELO FINAL

```
reg.mco10 = lm(Tasa_Paro~GP_EDUC_PC+GP_FOMENTO_PC + 'Var_Pobla_menor de
25'+ Coste_laboral+log(PIBPC)+IPC)
```

```
summary(reg.mco10)
```

```
#ANOVA
```

```
anova(reg.mco10)
```

```
#HISTOGRAMA DE RESIDUOS
```

```
hist(resid(reg.mco10))
```

```
#TEST SHAPIRO
```

```
shapiro.test(reg.mco10 $ residuals)\
```

```
#GRÁFICOS
```

```
plot(fitted.values(reg.mco10), #extraer los valores ajustados para Y  
resid(reg.mco10)) #extraer el valor de los residuos para cada caso X  
abline(h = 0, lty = 2) #add línea horizontal en 0
```

```
par(mfrow = c(2, 2))
```

```
plot(reg.mco10)
```

```
#REGRESIÓN CON EFECTOS FIJOS
```

```
reg_fijos=plm(Tasa_Paro~GP_EDUC_PC+GP_FOMENTO_PC + 'Var_Pobla_menor de 25'  
+ Coste_laboral+log(PIBPC)+IPC, data = mad, model = "within",  
index = c("cc.AA.", "Tiempo"))  
summary(reg_fijos)
```

```
reg_fijos2=lm(Tasa_Paro~GP_EDUC_PC+GP_FOMENTO_PC + 'Var_Pobla_menor de 25'  
+ Coste_laboral+log(PIBPC)+IPC + factor(CC.AA.)  
+ factor(Tiempo))  
summary(reg_fijos2)
```

```
**reg_fijos3=lm(Tasa_Paro~GP_EDUC_PC+GP_FOMENTO_PC + 'Var_Pobla_menor de 25'  
'+ Coste_laboral+log(PIBPC)+IPC + factor(CC.AA.))  
summary(reg_fijos3)
```

#REGRESIÓN CON EFECTOS ALEATORIOS

```
reg_aleatorio=plm(Tasa_Paro~GP_EDUC_PC+GP_FOMENTO_PC + 'Var_Pobla_menor de 25'  
+ Coste_laboral+log(PIBPC)+IPC, data = mad, model = "random")  
summary(reg_aleatorio)
```

#POOLED VS EFECTOS FIJOS

```
pFtest(reg_fijos, reg.mco10)
```

#TEST DE HAUSMAN

```
phtest(reg_fijos, reg_aleatorio)
```