

GRADO: Administración y Dirección de Empresas

Curso 2022/23

LA INCIDENCIA DEL COVID-19 EN LA TASA DE OCUPACIÓN DE LA COMUNIDAD AUTÓNOMA VASCA: ANÁLISIS ECONOMÉTRICO

Autora: Izaro García Lomas

Directores: Jorge Virto Moreno y Ainhoa Oguiza Tovar

Bilbao, a 13 de febrero de 2023



ÍNDICE DE CONTENIDO

1.	INTRODUCCIÓN Y MOTIVACIÓN	7
1.1	La tasa de ocupación	8
1.2	COVID - 19	9
2.	ENCUESTA DE POBLACIÓN ACTIVA	12
2.1	Datos reflejados en la EPA	13
2.2	Elaboración de la base de datos.....	16
3.	ANÁLISIS DESCRIPTIVO	19
3.1	Variables exógenas.....	19
3.1.1	Sexo	19
3.1.2	Edad.....	20
3.1.3	Nivel de formación	24
3.2	Variable endógena	26
3.2.1	Tasa de ocupación.....	26
4.	METODOLOGÍA.....	28
4.1	Modelo especificado	28
4.1.1	Variable endógena:	28
4.1.2	Variables exógenas:.....	28
4.2	Hipótesis.....	30
4.3	Contrastes de hipótesis.....	31
4.3.1	Hipótesis simples.....	32
4.3.2	Hipótesis múltiples.....	32
5.	MODELOS ESTIMADOS.....	34



5.1	Modelo para los años 2005-2020.....	34
5.1.1	Estimaciones del modelo inicial	34
5.1.1	Modelo inicial para las variables relevantes	36
5.2	Modelo para los años 2008-2020.....	38
5.2.1	Especificación de la tendencia mediante variables ficticias.....	38
5.2.2	Especificación de la tendencia de forma cuadrática.....	39
5.2.3	Comparación de la especificación para los años 2008-2020	41
5.3	Modelo para los años 2013-2020.....	42
5.3.1	Especificación de la tendencia mediante variables ficticias para los años 2013-2020 42	
5.3.2	Especificación de la tendencia de forma lineal	43
5.3.3	Comparación de la especificación para los años 2013-2020	44
5.4	Comparación de modelos estimados.....	45
5.5	Interpretación de los coeficientes.....	48
6.	CONCLUSIONES	51
7.	REFERENCIAS.....	53
8.	APÉNDICE	55
8.1	Modelo para el periodo 2008-2020 incluyendo todas las variables (incluso las no significativas).....	55
8.2	Modelo para el periodo 2013-2020 incluyendo todas las variables (incluso las no significativas).....	57

ÍNDICE DE FIGURAS

FIGURA 1: DISTRIBUCIÓN DE LA TASA DE OCUPACIÓN POR SEXO.....	19
FIGURA 2: DIFERENCIA DE PUNTOS PORCENTUALES DE LA TASA DE OCUPACIÓN DE LOS HOMBRES RESPECTO A LA DE LAS MUJERES.....	20



FIGURA 3: PORCENTAJE DE INDIVIDUOS POR GRUPOS DE EDAD	20
FIGURA 4: PIRÁMIDE POBLACIONAL DE LOS AÑOS 1991 Y 2021	21
FIGURA 5: DISTRIBUCIÓN DE TASA DE OCUPACIÓN POR EDAD	22
FIGURA 6: TASA DE OCUPACIÓN POR EDADES A LO LARGO DEL TIEMPO	23
FIGURA 7: PORCENTAJE DE INDIVIDUOS POR CADA NIVEL DE ESTUDIOS A LO LARGO DE LOS AÑOS	24
FIGURA 8: DISTRIBUCIÓN DE LA TASA DE OCUPACIÓN POR NIVEL DE ESTUDIOS	25
FIGURA 9: TASA DE OCUPACIÓN PARA CADA NIVEL DE FORMACIÓN	26
FIGURA 10: EVOLUCIÓN TEMPORAL DE LA TASA DE OCUPACIÓN (2005-2020)	26
FIGURA 11: ESTIMACIÓN TEMPORAL DE LA TASA DE OCUPACIÓN EN EL MODELO INICIAL	37
FIGURA 12: ESTIMACIÓN DE LA EVOLUCIÓN TEMPORAL 2008-2009: VARIABLES FICTICIAS VS TENDENCIA CUADRÁTICA.....	41
FIGURA 13: ESTIMACIÓN DE LA EVOLUCIÓN TEMPORAL: 2013-2020: VARIABLES FICTICIAS VS TENDENCIA LINEAL.....	44
FIGURA 14: COMPARACIÓN DE LOS DIFERENTES MODELOS ESTIMADOS	47

ÍNDICE DE TABLAS

TABLA 1: VARIABLES DE LA EPA QUE SE UTILIZARÁN PARA ESTE ANÁLISIS	15
TABLA 2: MUESTRA DE LA INFORMACIÓN PRESENTE EN LOS ARCHIVOS DE MICRODATOS	16
TABLA 3: ESTADÍSTICOS PRINCIPALES	27
TABLA 4: RESULTADO DE LAS ESTIMACIONES DEL MODELO INICIAL.....	34
TABLA 5: RESULTADO DE LAS ESTIMACIONES DEL MODELO INICIAL PARA LAS VARIABLES SIGNIFICATIVAS	36
TABLA 6: RESULTADO DE LAS ESTIMACIONES DEL MODELO CON VARIABLES FICTICIAS PARA LOS AÑOS 2008-2020 PARA LAS VARIABLES SIGNIFICATIVAS.....	39
TABLA 7: RESULTADO DE LAS ESTIMACIONES DEL MODELO CUADRÁTICO PARA LOS AÑOS 2008-2020 PARA LAS VARIABLES SIGNIFICATIVAS	40
TABLA 8: RESULTADO DE LAS ESTIMACIONES DEL MODELO CON VARIABLES FICTICIAS PARA LOS AÑOS 20013-2020 PARA LAS VARIABLES SIGNIFICATIVAS.....	43



TABLA 9: RESULTADO DE LAS ESTIMACIONES DEL MODELO LINEAL PARA LOS AÑOS 20013-2020 PARA LAS VARIABLES REPRESENTATIVAS	44
TABLA 10: COMPARACIÓN DE LOS DIFERENTES MODELOS	45
TABLA 12: RESULTADO DE LAS ESTIMACIONES DEL MODELO PARA LOS AÑOS 2008-2020 PARA TODAS LAS VARIABLES	55
TABLA 13: RESULTADO DE LAS ESTIMACIONES DEL MODELO CUADRÁTICO PARA LOS AÑOS 2008-2020 PARA TODAS LAS VARIABLES	56
TABLA 14: RESULTADO DE LAS ESTIMACIONES DEL MODELO CON VARIABLES FICTICIAS PARA LOS AÑOS 20013-2020 PARA TODAS LAS VARIABLES	57
TABLA 15: RESULTADO DE LAS ESTIMACIONES DEL MODELO LINEAL PARA LOS AÑOS 2013-2020 PARA TODAS LAS VARIABLES	58

RESUMEN

El objetivo principal del presente trabajo es tratar de estimar el efecto del COVID en la tasa de ocupación de la CAV. Para ello se ha usado un pseudopanel de datos calculado a partir de los microdatos del INE. Para poder estimar el efecto del COVID se modeliza la evolución temporal de la tasa de ocupación mediante varios modelos, uno con tendencia cuadrática y otro lineal. Ambos modelos se han estimado por MCO con la corrección de Arellano a la hora de estimar las varianzas de los estimadores. Se considera que el modelo lineal se ajustará mejor a la realidad y los resultados obtenidos a través de GRETL son los siguientes: se estima que el COVID generó un descenso de la tasa de ocupación del 1,3% en el año 2019 y de un 4,5% en 2020 en la CAV.

ABSTRACT

The main objective of this work is to try to estimate the effect of the COVID on the occupancy rate of the Basque Autonomous Community. In order to estimate the effect of the COVID it has been created a pseudo-panel of data using the microdatas of the INE and later the temporal evolution of the occupancy rate has been modeled using two different models, one with a quadratic trend and the other a lineal one. Both models have been estimated by ordinary least squares (OLS) with the Arellano variances corrections. The lineal model will be better adjusted to the reality than the quadratic ones, and the results obtained through GRETL are the following ones: it is estimated that COVID generated a decrease in the occupancy rate of 1.3% in 2019 and of 4.5% in 2020 in the Basque Autonomous Community.

LABURPENA

Lan honen helburu nagusia COVID-19ak EAEko okupazio-tasan eragindako eragina kalkulatzeko da. Horretarako INEk eskainitako datuetatik abiatuta datu pseudo-panel bat sortu da. Pandemiaren eragina kalkulatu ahal izateko okupazio-tasaren denbora-bilakaera modelizatu egin da, hainbat eredu ekonometriko proposatuz, bata joera koadratikoarekin eta bestea joera lineal bat jarraituz. Bi ereduak KTM bidez estimatu dira, Arellanok proposatutako zuzenketak erabiliz bariantzak kalkulatzeko orduan. Eredu lineala errealitateko parabolikoa baino hobeto egokitzen da, eredu parabolikoa COVID-aren eragina gainestimatu du eta. Eredua GRETL bidez estimatu ondoren, esan dezakegu 2019an okupazio tasak %1,3ko murrizketa jaso zuela pandemia dela eta, eta hurrengo urtean, 2020an, murrizketa hori %4,5koa izan zela.

1. INTRODUCCIÓN Y MOTIVACIÓN

Primero la crisis sanitaria y más tarde la crisis económica causadas por la pandemia del COVID-19 han tenido influencia en la sociedad, en el desarrollo tecnológico, la política y cómo no, en la economía. A lo largo de este trabajo se analizará la influencia que ha tenido la pandemia en la tasa de ocupación de la Comunidad Autónoma Vasca (CAV).

A finales de diciembre de 2019, China informó sobre los primeros casos de pacientes con una neumonía con etiología desconocida, que más tarde se conocería como COVID-19. Esos casos puntuales comenzaron a expandirse primero en China, y poco a poco por Asia, EE. UU. y Europa, pero no fue hasta el 11 de marzo cuando la Organización Mundial de la Salud declaró la pandemia de COVID-19, lo que supuso que se tomaran diversas medidas en todo el mundo para frenar su expansión (Dirección de Economía y Planificación, 2020).

En España el Gobierno de Pedro Sánchez declaró el estado de alarma durante quince días a partir del 14 de marzo de 2020. Dicho decreto dio lugar a un confinamiento de la población que se mantuvo hasta el 11 de mayo del mismo año. A raíz del estado de alarma, muchos negocios y establecimientos se suspendieron temporalmente (a excepción de los establecimientos comerciales de primera necesidad). En consecuencia, muchos empleos fueron temporalmente “innecesarios”, ya que todo tipo de museos, bibliotecas, monumentos, restauración, hostelería e instalaciones deportivas quedaron cerrados de cara al público hasta el 14 de abril del mismo año. Pese a todas esas medidas, España, junto a Italia, soportó una de las incidencias acumuladas más elevadas de Europa, con unos 1.000 fallecidos diarios por coronavirus durante los primeros meses de confinamiento (Dirección de Economía y Planificación, 2020 y Real Decreto Ley 463/2020).

Al comienzo de la crisis sanitaria, la tasa de desempleo sufrió un incremento importante. Para poder frenar el aumento del paro se llevaron a cabo distintas medidas, como por ejemplo ayudas económicas a distintas empresas o los Expedientes de Regulaciones Temporales de Empleo (ERTE). Estas medidas fueron establecidas a través del Real Decreto Ley 8/2020, donde también se prorrogó el subsidio por desempleo y se dio prioridad al trabajo no presencial siempre y cuando fuera posible (Real Decreto-ley 8/2020). A través de estas medidas se protegió la economía de 3,6 millones de personas (un 20% de las personas afiliadas a la Seguridad Social en abril de 2020), y se intentó evitar un despido masivo de trabajadores para frenar la crisis económica. En el caso del País Vasco, más de 220.000 personas estuvieron bajo un ERTE durante 2020. A pesar de estas medidas, la evolución del mercado laboral a nivel estatal fue bastante negativa, ya que quienes estaban en búsqueda de empleo (en su mayoría jóvenes), personas en desempleo de larga duración, o quienes perdieron su empleo temporal, se encontraron con grandes dificultades para encontrar trabajo y además no pudieron acogerse a los ERTEs.

Por otro lado, hay que tener en cuenta que durante el confinamiento aumentó la tasa de inactividad, ya que muchas de las personas que perdieron su trabajo no pudieron incorporarse al



mundo laboral ni realizar una búsqueda activa de empleo. En este caso se encontrarían quienes durante el confinamiento tuvieron algún tipo de enfermedad (muchos de ellos con COVID) o quienes tuvieron que realizar cuidados de personas mayores, niños o personas con discapacidad (Ruesga & Viñas, 2020 y Dirección de Economía y Planificación, 2020).

A nivel estatal, las comunidades autónomas que implantaron medidas más restrictivas tuvieron una incidencia menor, pero una tasa de desempleo mayor (Gutiérrez & Moral-Benito, 2020). Durante la COVID-19 hubo una dicotomía entre proteger la economía o la salud pública. Otros factores como por ejemplo la densidad de la población, el nivel de envejecimiento poblacional o el clima también fueron influyentes en la incidencia de casos coronavirus (Gutiérrez & Moral-Benito, 2020).

El primer caso de coronavirus detectado en el país vasco fue el 29 de febrero. En cuestión de días, aumentó exponencialmente el número de casos de COVID-19 y durante toda la pandemia el País Vasco tuvo una tasa de incidencia elevada (Yeregui & Guisasola, 2021). El empleo vasco durante 2020 se redujo un 8,1%, lo que supuso una reducción de 80.000 puestos de trabajo menos respecto al año anterior. Hay que tener en cuenta, que durante ese periodo el Producto Interior Bruto (PIB) vasco tuvo una disminución de 9,5%, siendo el sector industrial el más afectado (una disminución del 11,8%), y el sector de servicios públicos el único que creció, con un aumento del 1%. Entre dichos servicios públicos se encuentran la sanidad, administración pública y educación, servicios que fueron imprescindibles para el control de la pandemia (Dirección de Economía y Planificación, 2020).

1.1 La tasa de ocupación

Para dar comienzo al análisis de la incidencia de la pandemia en la tasa de ocupación vasca, se definirá lo que es la tasa de ocupación. La tasa de ocupación, también conocida como tasa de empleo, se define como el cociente entre el número total de ocupados y la población en edad de trabajar.

La población en edad de trabajar será entendida como el rango de edad entre los dieciséis años (edad mínima para acceder al mundo laboral) en adelante. En este rango de personas en edad de trabajar hay quienes establecen un rango máximo de edad de sesenta y cuatro años y hay quienes no definen ningún máximo. Aunque estadísticas como la encuesta de población activa (EPA) recojan datos de individuos de edades superiores a los sesenta y cuatro años y el único requisito de edad que establecen para analizar el empleo es que el individuo sea mayor de dieciséis años, para este análisis se utilizará dicho máximo, ya que los datos de los individuos mayores de sesenta y cuatro años tienden a tener tasas de ocupación menores debido a la jubilación.

Por otro lado, la población económicamente activa (PAE) se define como la población que o bien aporta su trabajo para producir bienes y servicios económicos o bien las personas que carecen

de empleo en ese momento concreto pero lo buscan y están disponibles para incorporarse a un empleo (Greenwood, 1999).

Entre los factores que inciden a la tasa de ocupación cabe destacar la edad media poblacional, ya que las tasas de desempleo no se reparten de manera homogénea a lo largo de las distintas edades, siendo las personas más jóvenes, y las de más edad, quienes sufren una tasa de desempleo mayor. Lo mismo ocurre con otros factores como podrían ser la nacionalidad o el nivel de estudios. Existen segmentos de población que tienden a tener empleos más precarios y una tasa de temporalidad mayor, como son ser los jóvenes, los extranjeros o quienes no disponen de cierto nivel de estudios. Por ello, no solo sufren unas tasas de desempleo más altas, sino que padecen en mayor medida las consecuencias de las crisis económicas, ya que los puestos de trabajo que ocupan son los primeros en desaparecer en momentos de recesión económica (Medina, 2004 y Rocha, 2012).

Los factores que afectan a la tasa de ocupación van variando con el paso del tiempo y no siempre tienen la misma influencia. Una modificación en la cobertura de los subsidios puede generar un cambio en la tasa de ocupación, por ejemplo, ya que los individuos podrán optar por acogerse a los subsidios y no incorporarse a la vida laboral (Roncaglia, 2006).

1.2 COVID - 19

El objetivo de este trabajo es el análisis del modo en el que ha influido la crisis económica del COVID en la tasa de ocupación de la CAV, teniendo en cuenta las crisis económicas no afectan de manera homogénea a toda la población. Además, el Gobierno puso en marcha los ERTes con el fin de paliar los efectos del parón económico provocado por el COVID pero estas medidas no alcanzaron a toda la población de manera homogénea, y así el efecto de la crisis no fue el mismo en aquellas personas que pudieron acogerse o no a dichas medidas (Mackay-Castro et al., 2020).

Se deberá tener en cuenta que antes de que comenzara la pandemia del COVID, España ya registraba la segunda tasa de paro más alta de Europa, solo superada por Grecia, y esta situación era aún peor en el mercado laboral juvenil (menores de 25 años). Esto obligó a los jóvenes a aceptar trabajos más precarios, pese a tener un alto nivel de estudios, y por ello, sufrieran en mayor medida los efectos de la crisis. El mercado laboral antes de la pandemia era el resultado de la crisis económico-financiera del 2009 y de su gestión, ya que el empleo creado para paliar los efectos de la crisis, fue de menor calidad, aumentando la precariedad laboral del mercado (Coll & López, 2020).

Según las cifras facilitadas por la Seguridad Social respecto a la afiliación, podemos observar que a partir del 11 de marzo, dos días antes de la declaración del estado de alarma, en un periodo de un mes 898.822 afiliados perdieron su empleo y de estos 613.250 tenían un contrato temporal (70% de la destrucción), pese a que en España los contratos temporales masculinos representen tan solo un 11,8% y los femeninos un 12,3%. Los empleados con este tipo de contratos soportaron una pérdida de empleo relativa mucho mayor porque son estos trabajos los primeros



en desaparecer en épocas de crisis económicas (Instituto Nacional de Estadística, 2021 y Ruesga & Viñas, 2020).

El empleo juvenil se define como el empleo de la población entre dieciséis y veinticuatro años. Cabe destacar que este segmento de la población suele tener cifras de actividad menores, ya que muchos jóvenes están en periodo de formación, por lo que no están en búsqueda activa de empleo. Pese a tener una tasa de actividad menor, su tasa de desempleo es de las más altas, con una tasa de paro antes de la pandemia¹ de un 33,53%, y una tasa de temporalidad de un 70%. En octubre de 2020, la tasa de desempleo juvenil en España alcanzó un 38,9% para los hombres y un 42,2% para las mujeres, según la Organización para la Cooperación y el desarrollo Económico (OCDE). Estos datos fueron los más elevados de Europa (la media europea² se situaba en un 14,9% para los hombres y en un 13,4% para las mujeres), y aunque la situación antes de la pandemia era preocupante, la crisis del COVID-19 no hizo más que empeorar las cifras (López, 2020).

Respecto a la brecha de género, cabe destacar que sufrió un aumento tras la pandemia. La inactividad creció en mayor medida para las mujeres, ya que tienden a encargarse de los cuidados de mayores, niños o personas con discapacidad. Por ello, durante la cuarentena se vieron en la necesidad de quedarse atendiendo estos cuidados. Además, son las mujeres quienes tienen una mayor tasa de desempleo de larga duración y de contratos temporales, factores que generan un mayor impacto de los efectos de la crisis (Ruesga & Viñas, 2020). Se deberá tener en cuenta, que antes de la pandemia de COVID-19, la tasa de desempleo de las mujeres era del 15,5% y la de los hombres del 12,2%. En cambio, tras la crisis³, la tasa de desempleo de las mujeres alcanzó un 17,4% (un aumento de 1,8 puntos porcentuales) mientras que la tasa de desempleo de los hombres fue de un 13,4% (aumento de 1,2 puntos porcentuales). Por ello, se puede afirmar que antes de la pandemia ya existía una brecha de género en el mercado laboral, y esta brecha se ha agrandado como consecuencia COVID-19 (Gómez et al., 2021).

¹ A finales de 2018

² Información respecto a 28 países.

³³ Datos correspondientes al segundo trimestre de 2021.



Otro factor a tener en cuenta es a nacionalidad. Mientras que la disminución de la tasa de empleo entre enero y junio de 2020 para los españoles fue de un 6,1 %, para los extranjeros supuso un 11,4 %. De hecho, solo de enero a marzo, el paro aumentó en un 0,4% entre los españoles, mientras que el desempleo extranjero aumentó en un 17,7%. A mediados de 2020 la tasa de desempleo entre los extranjeros fue de un 24,9%, y la de los españoles tan solo de un 11,4%. En términos interanuales, septiembre de 2019 frente a septiembre de 2020, el paro entre las personas extranjeras ascendió un 44%, y el de la media nacional un 23%, casi la mitad. Esta diferencia, al igual que lo que ocurre con las mujeres o los jóvenes, se debe en gran medida a la alta temporalidad de los contratos, menor antigüedad laboral y mayor precariedad. Además, los sectores con más trabajadores de origen extranjero fueron los más golpeados por la pandemia; por ejemplo, las trabajadoras del hogar, donde un 42% de las afiliadas a la Seguridad Social de este sector son de nacionalidad extranjera (Mahía, 2021).

El último factor a analizar es el nivel de estudios de las personas trabajadoras. En el año 2020 la tasa de empleo entre las mujeres con un nivel de educación superior (nivel 5-8) era de un 75,2%, mientras que la tasa de empleo de las mujeres con estudios hasta la segunda etapa de educación secundaria (nivel 3-4) era de un 55,7%, y la tasa entre las mujeres con un nivel de estudios de preescolar, primaria o primera etapa de educación secundaria (nivel 0-2) era de un 44,3%. Algo similar ocurre con los hombres. Los segmentos con mayor nivel de estudios, tienden a tener mayores tasas de empleo. La tasa de desempleo entre los hombres con nivel de educación de preescolar, primaria o primera etapa de educación secundaria (nivel 0-2) en 2020 fue de un 18,7% mientras que la de quienes tenían un nivel de educación superior (nivel 5-8) baja hasta un 8,9% (Instituto Nacional de Estadística, 2021).

Durante el confinamiento, se paralizaron la mayoría de los empleos, excepto los trabajos considerados esenciales y los que se podían realizar de manera telemática. Hay que tener en cuenta, que no todos los trabajos pueden realizarse de manera telemática, y en caso de poder hacerlo, se necesita una inversión en equipos informáticos. En gran medida, los empleos menos cualificados son más de carácter presencial o manual, y, por ende, no se pueden llevar a cabo de manera no presencial. En cambio, los trabajos que requieren de un mayor nivel de cualificación se pueden ejecutar de manera telemática. Por ello, fueron las personas que tenían mejores cualificaciones, las que pudieron realizar su trabajo sin interrupción durante el confinamiento y por lo tanto continuaron con su empleo sin necesidad de estar en paro o ERTE. Por otro lado, quienes contaban con un nivel de estudios menor, o bien continuaron con su empleo de manera presencial (en caso de ser considerado un servicio esencial) o se vieron con la obligación de dejar de trabajar.

2. ENCUESTA DE POBLACIÓN ACTIVA

Para llevar a cabo el análisis se utilizará la Encuesta de Población Activa, también conocida como EPA. La EPA es una encuesta facilitada por el Instituto Nacional de Estadística (INE), que se realiza trimestralmente mediante el análisis de la población residente en viviendas familiares a nivel nacional y cuyo tamaño muestral es el siguiente: 3.822 secciones censales, alrededor de 65.000 viviendas y 160.000 personas. La EPA es una encuesta continua, es decir, se lleva a cabo durante todo el año de manera homogénea. Las muestras son recogidas en viviendas familiares, entendiéndose como viviendas familiares “residencias en las que se comparte un presupuesto común entre los miembros” (Instituto Nacional de Estadística, 2021).

La Encuesta de Población Activa recoge datos sobre la relación entre la población y el mercado laboral. Gracias a esta encuesta es posible realizar comparaciones en el tiempo, entre distintas comunidades autónomas e incluso entre distintos países, ya que la encuesta se realiza acorde a los estándares establecidos a nivel internacional (Instituto Nacional de Estadística, 2021).

Hay que tener en cuenta que, al referirse a una encuesta de muestreo, puede haber error de muestreo, además de falta de detalle en algunos aspectos. Al no tratarse de una encuesta realizada a toda la población, es posible que no se recojan ciertas realidades del mundo laboral y que ciertos datos queden bastante generalizados (Instituto Nacional de Estadística, 2021). Pese a que la encuesta tenga cierto margen de error, sigue tratándose de una encuesta de gran fiabilidad y estabilidad. De hecho, es una encuesta de gran importancia, por su alta cobertura representativa de la población, y por su gran obtención de datos a lo largo del tiempo, ya que se efectúa desde 1964 (Villagarcía, 1995).

La primera EPA fue publicada por el INE en el segundo trimestre de 1964 y con el fin de adecuarse a las realidades del mundo laboral ha ido experimentando diversos cambios en su metodología de recogida de datos. Por ejemplo, desde 1969 hasta 1975 la EPA fue realizada de manera semestral, en lugar de trimestral. En 1976 se añadió la información respecto al paro, y más tarde en 1987 se añadió información sobre los nuevos métodos de contratación, ya que en dicho año se llevó a cabo una reforma del Estatuto de los trabajadores. A través de esta segunda medida, se logró adaptar la EPA a los estándares establecidos por la Oficina Estadística de la Unión Europea (EUROSTAT) y por ello, es posible realizar comparaciones entre los distintos países de la Comunidad. Ese cambio fue llevado a cabo por recomendación de la Conferencia Internacional de Estadísticos del Trabajo de 1982, debido a la entrada a la Unión Europea de España en 1996 (Gómez et al., 1994). Posteriormente, en 1992, se llevó a cabo otra reforma en la metodología de la EPA con intención de obtener más información sobre las condiciones laborales de los trabajadores y sobre las principales características demográficas de la población extranjera residente en España, además de reunir información sobre la educación recibida de la población. En el año 2002 se volvieron a realizar cambios respecto a la definición del paro (Gómez & García-Perea, 1994 y Torres, 2020).

El cambio más notorio se llevó a cabo en el año 2005, cuando se introdujo un nuevo cuestionario, más comprensible, y adaptado a las tecnologías del momento. Dicho cuestionario es la base del cuestionario actual, y utiliza encuestas telefónicas asistidas por ordenador. En ese mismo año también se adaptaron los periodos de referencia, dando lugar a variables de submuestra anual; ya que hasta entonces todas las variables eran de carácter trimestral. Debido a los grandes cambios que se dieron en 2005, los datos anteriores a ese año no son tan homogéneos y dificultan el correcto análisis y seguimiento de estos, por ello en este análisis solo se utilizarán datos a partir del 2005, puesto que se puede considerar que los años que han pasado años son suficientes como para recolectar los datos necesarios. El último cambio llevado a cabo se realizó el segundo cuatrimestre del 2020, a través del cual se han añadido preguntas extra a raíz de la pandemia del COVID-19 y hoy en día en la EPA se recogen datos sobre el teletrabajo, causa de despido e información sobre los horarios de trabajo (Torres, 2020 e Instituto Nacional de Estadística, 2021).

La EPA no es la única metodología utilizada para la recogida de datos del mercado laboral, y existen otros métodos para el análisis del empleo, como, por ejemplo: censos de población, encuestas de salarios, datos del paro registrado o afiliación a la Seguridad Social. La EPA dispone de diversas ventajas respecto al resto de bases de datos del mercado laboral y entre dichas ventajas están: los entrevistadores son especializados en este campo, los resultados son rápidos, ya que al ser una encuesta de muestreo no es necesario recoger datos de toda la población y las definiciones son lo más homogéneas posibles a lo largo de los años, lo que facilita el análisis del empleo a lo largo de los años (Torres, 2020).

2.1 Datos reflejados en la EPA

Entre los datos reflejados en la EPA para cada individuo se pueden destacar catorce categorías de datos recogidos: datos de control, demográficos, formación y nivel de estudios, trabajo en la semana de referencia, datos del empleo principal, del segundo empleo, deseo de trabajar más horas, búsqueda de empleo, experiencia profesional, inscripción en la oficina de empleo pública, información sobre situaciones diversas, días de ausencia, información sobre si se trabajó el año pasado y variables derivadas (Instituto Nacional de Estadística, 2021).

Para este trabajo que intenta explicar la influencia de la pandemia en el mercado laboral de la CAV, los datos que resultan más determinantes son: el sexo del individuo, edad, origen, nivel de estudios, la clasificación de los entrevistados en relación con la actividad económica según los criterios de la Oficina Internacional del trabajo (OIT) (ocupado, parado o inactivo) o si anteriormente tenía un trabajo con contrato temporal (Dirección de Economía y Planificación, 2020).

A continuación, se enumeran las variables de la EPA que se utilizarán para este análisis:

1. Sexo (Sexo1):

Variable clasificada como hombre y mujer: V (hombres) + M (mujeres).

2. Edad (Edad1):

Los datos están presentados en grupos quinquenales de años cumplidos: 00 (de 0 a 4 años), 05 (entre 5 y 9 años), 10 (entre 10 y 15 años), 16 (entre 16 y 19 años), 20 (entre 20 y 24 años), 25 (entre 25 y 29 años), 30 (entre 30 y 34 años), 35 (entre 35 y 39 años), 40 (entre 40 y 44 años), 45 (entre 45 y 49 años), 50 (entre 50 y 54 años), 55 (entre 55 y 59 años) y 60 (mayores de 60 años).

No se utilizarán datos referidos a edades inferiores a 25 años, ya que esa franja de edad tiende a presentar tasas de ocupación menores debido a que en buena medida continua con sus estudios. Tampoco se utilizarán datos de individuos mayores de 65 años. Por ello se dejarán a un lado los grupos de edad 00, 05, 10 y 16.

3. Nivel de formación (Nforma):

Variable que clasifica los estudios realizados: AN (analfabetos) + P1 (educación primaria incompleta) + P2 (educación primaria) + S1 (primera etapa de educación secundaria) + SG (segunda etapa educación secundaria, orientación general) + SP (segunda etapa educación secundaria, etapa profesional) + SU (educación superior).

Para este análisis se agregarán estos datos en las siguientes categorías: Obligatoria (O) (analfabetos, educación primaria incompleta, educación primaria y primera etapa de educación secundaria) + Orientación General (S) (segunda etapa de educación secundaria) + Orientación Profesional (P) (segunda etapa de educación secundaria) + Educación Superior (U) (educación superior)

4. Ocupación (AOI):

Clasificación de los entrevistados en relación con la actividad según la OIT (AOI). En la EPA están reflejados de la siguiente manera: 03 (ocupados subempleados por insuficiencia de horas) + 04 (resto de ocupados) + 05 (parados que buscan su primer empleo) + 06 (parados que han trabajado antes) + 07 (inactivos desanimados) + 08 (activos potenciales) + 09 (resto de inactivos).

En este análisis se agregarán los datos de esta manera: O (ocupados) + NO (no ocupados), entendiendo como la población ocupada a *“las personas de 16 o más años que durante la semana de referencia han estado trabajando durante al menos una hora a cambio de una retribución en dinero o especie, o quienes teniendo trabajo han estado temporalmente ausentes del mismo por enfermedad, vacaciones, etcétera”* según los criterios del INE.

Los ocupados pueden ser trabajadores por cuenta propia o asalariados, y pueden clasificarse en ocupados a tiempo completo (jornada habitual semanal superior a 30 horas) o a tiempo parcial

(jornada habitual semanal inferior a 35 horas), pero en este análisis no se hará esta distinción (Instituto Nacional de Estadística, 2021).

En la siguiente tabla se agregarán los códigos de las variables de manera que sean de mayor utilidad. La columna número tres – *Códigos EPA* – clasificación de las variables como aparecen en la EPA, mientras que la columna número cuatro – *Niveles TFG* – clasificación de las variables en este TFG, en algún caso agregando clases similares para facilitar el análisis.

Tabla 1: Variables de la EPA que se utilizarán para este análisis

<i>Variables</i>	<i>Definición</i>	<i>Códigos EPA</i>	<i>Niveles TFG</i>
Sexo 1	Sexo del individuo	1- Hombre	V - Hombre
		6- Mujer	M - Mujer
Edad 1	Edad del individuo. Los datos están presentados en grupos quinquenales de años cumplidos.	25 - Entre 25 y 29 años	1 - Entre 25 y 29 años
		30 - Entre 30 y 34 años	2 - Entre 30 y 34 años
		35 - Entre 35 y 39 años	3 - Entre 35 y 39 años
		40 - Entre 40 y 44 años	4 - Entre 40 y 44 años
		45 - Entre 45 y 49 años	5 - Entre 45 y 49 años
		50 - Entre 50 y 54 años	6 - Entre 50 y 54 años
		55 - Entre 55 y 59 años	7 - Entre 55 y 59 años
		60- Entre 60 y 64 años	8 - Entre 60 y 64 años
Nforma	Estudios completados del individuo	AN - analfabetos	
		P1 - educación primaria incompleta	O - Obligatoria
		P2 - educación primaria	
		S1 - primera etapa de educación secundaria	
		SG - segunda etapa educación secundaria, orientación general	S - Orientación general
		SP - segunda etapa educación secundaria, etapa profesional	P - orientación profesional
		SU - educación superior	U - Educación Superior
AOI	Clasificación de los entrevistados por relación con la actividad	03 - ocupados subempleados por insuficiencia de horas	O - Ocupados
		04 - resto de ocupados	
		05 - parados que buscan su primer empleo	N - No Ocupados

06 - parados que han trabajado antes

07 - inactivos desanimados

08 - activos potenciales

09 - resto de inactivos

Fuente: elaboración propia

2.2 Elaboración de la base de datos

La información facilitada por el INE a través de la EPA recoge de manera homogénea una gran cantidad de datos disponibles para cada uno de los trimestres entre 2005 y el primer trimestre de 2022. Algunos de esos datos quedarán fuera de este análisis, por ejemplo, los datos de los individuos que no pertenecen a la CAV, o la información respecto al anterior empleo del individuo. Por ello, se deberá filtrar la información de la EPA dejando a un lado toda la información que no sea relevante para este análisis.

Además, la EPA refleja datos de ciertos individuos a lo largo del tiempo, pero para poder analizar las tasas de ocupación llevando a cabo un modelo lineal, será necesario que los datos utilizados estén agrupados en grupos de individuos en lugar de utilizar datos de individuos sueltos. Esto se debe a que, a la hora de calcular las tasas de ocupación, será necesario dividir el número de ocupados entre el total de individuos pero los datos individuales de los ficheros no permiten ese cálculo. Por ello los individuos se agruparán en pseudo-individuos respecto a un conjunto de características comunes: sexo, edad y nivel de estudios..

Al tratarse de pseudo-individuos en lugar de individuos, obtenemos una configuración similar a los datos de panel: el pseudo-panel. Los paneles sirven para poder analizar los datos respecto a una característica, el tiempo en este caso, con intención de seguir un patrón o comportamiento (Deaton, 1985 y Oguiza et al., 2012). De esta manera se analizarán las tasas de ocupación de cada uno de los pseudo-individuos a lo largo de los años, para poder observar su evolución y en qué medida han afectado cada una de las características de los individuos. De no agregar los individuos de esta manera, no sería posible seguir la evolución a lo largo del tiempo de cada pseudo-individuo, ya que los individuos encuestados en la EPA solo serán encuestados durante año y medio, y posteriormente serán remplazados por individuos de similares características.

Tabla 2: Muestra de la información presente en los archivos de microdatos

	CICLO	CCAA	EDAD	SEXO	NFORMA	AOI	FACTOREL
1	194	País Vasco	40-44 años	Varón	Educación superior	Resto de ocupados	19390
2	194	País Vasco	30- 34 años	Mujer	Educación superior	Ocupados subempleados por insuficiencia de horas	19390

3	194	País Vasco	35 - 39 años	Mujer	Educación superior	Resto de ocupados	25232
4	194	País Vasco	30 -34 años	Varón	Educación superior	Resto de ocupados	25232
5	194	País Vasco	60 - 64 años	Mujer	Educación superior	Resto de ocupados	24658
6	194	País Vasco	25 - 29 años	Mujer	Educación superior	Parados que han trabajado antes	24658
7	194	País Vasco	45 - 49 años	Mujer	Educación superior	Inactivos (resto de inactivos)	14401
8	194	País Vasco	45 - 49 años	Varón	Segunda etapa de educación secundaria, orientación general	Resto de ocupados	14401
9	194	País Vasco	16 - 19 años	Varón	Primera etapa de educación secundaria	Inactivos (resto de inactivos)	14401
10	194	País Vasco	50 - 54 años	Varón	Segunda etapa de educación secundaria, orientación general	Ocupados subempleados por insuficiencia de horas	24624

Fuente: Elaboración propia a partir de los microdatos de la EPA (INE)

Esta tabla es un ejemplo en el que se reflejan los diez primeros individuos del fichero de microdatos respecto a las variables seleccionadas en este análisis para el primer trimestre de 2021, individuos que más tarde que se agruparán para crear los pseudo-individuos para un análisis correcto. Para ello se enumerará el total de individuos respecto a cada una de las diferentes combinaciones de sexo, edad y formación. Posteriormente, para calcular la tasa de ocupación se analizará cuántos individuos de cada uno de los pseudo-individuos están ocupados. Este proceso de conteo se deberá llevar a cabo con todos los individuos que se analizarán, no solo con los diez primeros.

En la tabla 2, la primera columna enumera cada uno de los individuos, mientras que la columna del ciclo hace referencia al trimestre al que pertenecen esos datos, en este caso el ciclo 194 hace referencia al primer trimestre del 2021 (como ya se explicará más adelante, en este análisis se agrupará con 2020). Por otro lado, únicamente se usarán los datos de la CAV, por lo que todos los individuos analizados tendrán esta característica. Las variables de edad, sexo, nivel de formación y clasificación según la AOI hacen referencia a las características de cada pseudo-individuo que utilizaremos para calcular las tasas de ocupación. La variable *factorel* significa el



número proporcional al que representa cada individuo, para conocer el número exacto de individuos al que representa cada pseudo-individuo se deberá dividir el factoriel entre 100.

Estos pseudo-individuos se construirán en función de la variable sexo (V–M), edad (1–8) y nivel de formación (O, S, P, U) dando lugar a sesenta y cuatro tipos de pseudo-individuos ($2 \times 8 \times 4 = 64$). Estos pseudo-individuos formarán un pseudo-panel a lo largo de dieciséis años, por lo tanto, se calcularán 1.024 tasas de ocupación (una para cada pseudo-individuo cada año) ($64 \times 16 = 1.024$).

Por ejemplo, utilizando los datos del 2005 hay 112 mujeres del primer tramo de edad (25-29 años) con estudios obligatorios, entre ellas 71 están ocupadas y el resto no lo están. Así, la tasa de ocupación para el 2005 para este tipo de pseudo individuo (MO1) será del 63,39% ($71/112$). Este mismo proceso se ha llevado a cabo con todos los pseudo-individuos para cada año, obteniendo así las ya antes mencionadas 1.024 tasas de ocupación.

Por otro lado, se deberá tener en cuenta que se calculará una tasa de ocupación para cada pseudo-individuo por año, pero como el objetivo de este trabajo es analizar el efecto que ha tenido la pandemia en la tasa de ocupación de la CAV y las consecuencias del coronavirus empezaron en el segundo trimestre, será más correcto agrupar los trimestres de esta forma: como año 2020 se considerarán los 2º, 3º y 4º trimestres del 2020 y el 1º del 2021, y lo mismo con el resto de los años. Por ello, el 1º trimestre de 2020 pertenecerá al 2019. De esta forma, se consigue aislar el efecto del COVID en un solo año, ya que en el 1º trimestre de 2020 todavía no había afectado demasiado el virus, y todavía los datos del 4º trimestre de 2021 no están publicados en el INE (a fecha de 2021). Aun así, no se logrará aislar totalmente el efecto de la pandemia en un único año, ya que las primeras consecuencias del COVID comenzaron a aparecer en el mes de marzo de 2020, mes que contabilizará como parte del 2019.

3. ANÁLISIS DESCRIPTIVO

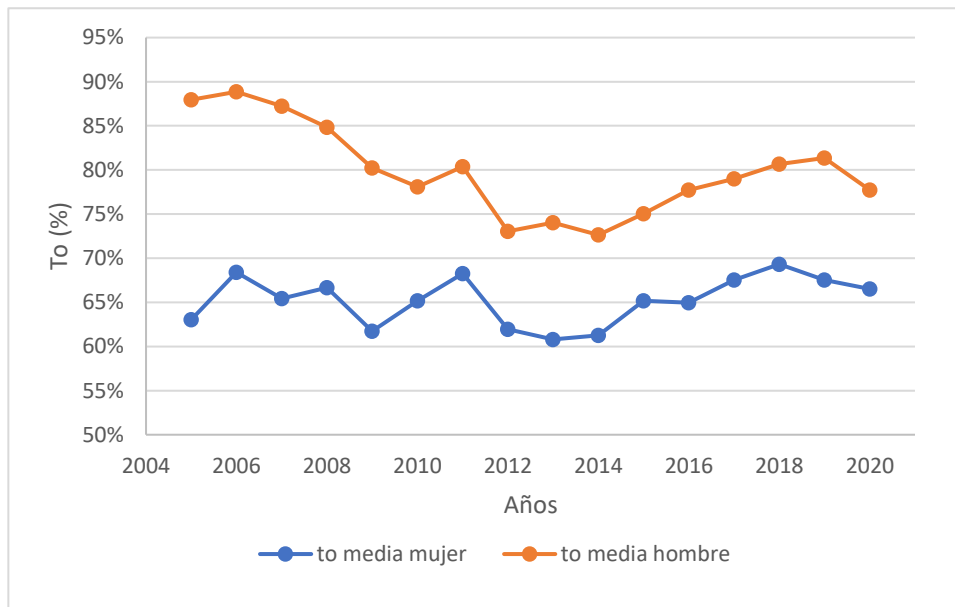
A continuación, se realizará un análisis descriptivo de las diferentes variables de este análisis.

3.1 Variables exógenas

3.1.1 Sexo

El número total de mujeres y de hombres reflejados en la EPA está bastante igualado. En los últimos años, las mujeres han contado con mayor representación en la EPA respecto a los hombres, pero esta diferencia corresponde a la realidad de la sociedad de la CAV, ya que en 2021 en la CAV residían 1.129.158 mujeres y 1.064.041 hombres (Euskal Estatistika Erakundea, 2022). Es decir, las mujeres representaban en 2021 el 51,48% de la población de la CAV.

Figura 1: Distribución de la tasa de ocupación por sexo

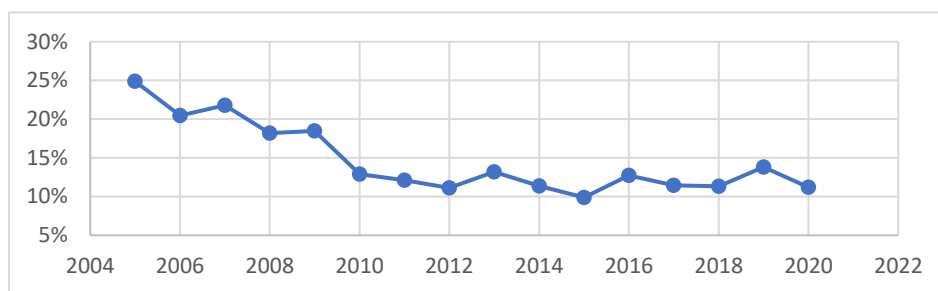


Fuente: Elaboración propia a partir de los datos de la EPA (INE)

Respecto al sexo, como puede observarse en la figura 1, las tasas de ocupación de ambos sexos suelen tener la misma evolución, pero los hombres tienden a tener mayores tasas de ocupación que las mujeres. Esto es un reflejo de la menor presencia de la mujer en el mercado laboral, debido a que tradicionalmente se han ocupado a los trabajos del hogar. De hecho, la tasa de ocupación media de las mujeres para los años analizados es de 63,29% mientras que la de los hombres es de 76,35%.

En la figura 1 se puede apreciar como las tasas de ocupación femeninas se mantienen mas constantes a lo largo de los años mientras que la de los varones sufren mayores cambios a lo largo de los años.

Figura 2: Diferencia de puntos porcentuales de la tasa de ocupación de los hombres respecto a la de las mujeres



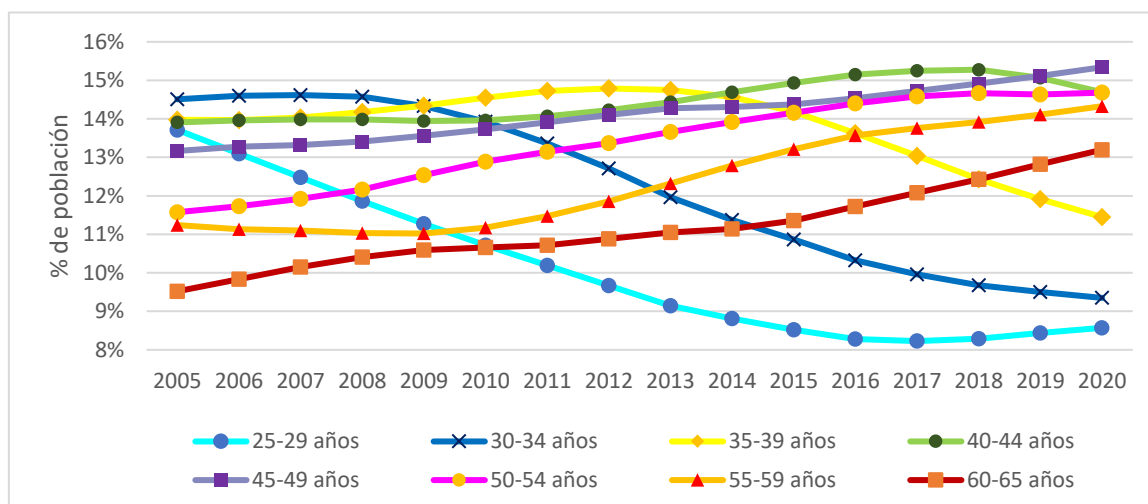
Fuente: Elaboración propia a partir de los datos de la EPA (INE)

Aunque las mujeres presentan tasas de ocupación más bajas que los varones, se percibe que la brecha de género a lo largo de los años va disminuyendo. Así, en 2007, la diferencia entre la tasa de ocupación de los hombres y la de las mujeres era de 21,71 puntos porcentuales, mientras que en 2020 la diferencia entre tasas es de 11,21 puntos porcentuales.

Esta brecha de género disminuyó especialmente tras la crisis del 2008, debido no a la mejora de la situación laboral de las mujeres, sino a una bajada significativa de la tasa de ocupación masculina, ya que las tasas de ocupación femeninas de antes de la crisis financiera del 2008, y las de antes de la pandemia son bastante similares mientras que las masculinas han pasado de tener valores superiores al 85% antes de la crisis, a estar alrededor de un 80% pre-covid. Además, aunque la brecha de género haya disminuido en 2020 respecto a 2005, en los últimos diez años se ha mantenido medianamente constante, con unas tasas de ocupación masculinas un 12% superiores al de las mujeres.

3.1.2 Edad

Figura 3: Porcentaje de individuos por grupos de edad

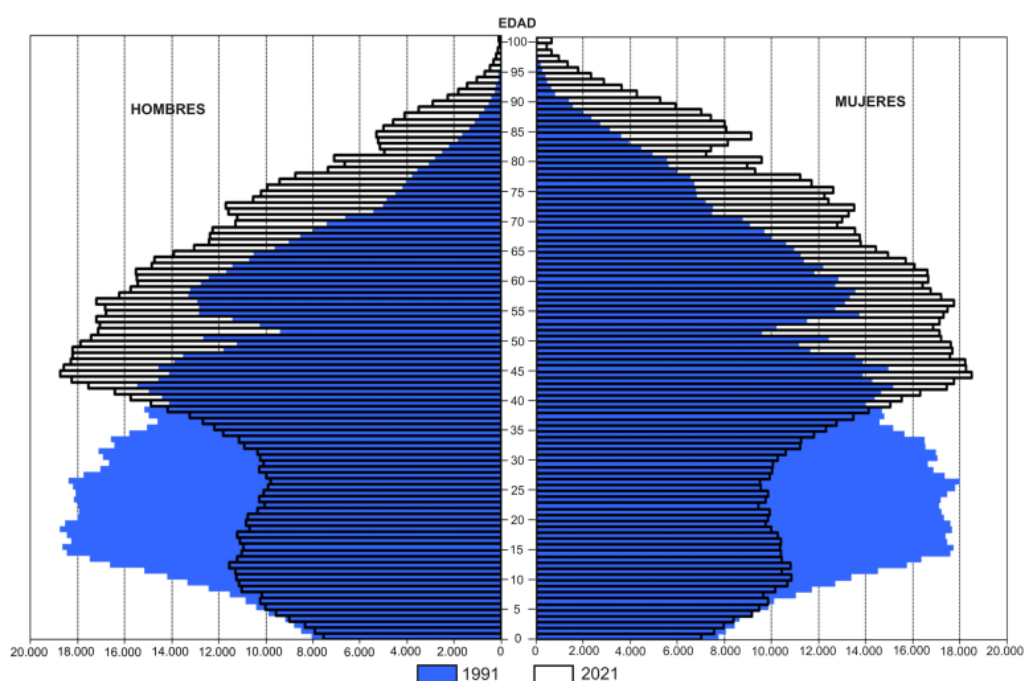


Fuente: Elaboración propia a partir de los datos de la EPA (INE)

Como se puede observar en la figura 3, a lo largo de estos años ha habido un descenso llamativo de la población más joven (edades entre los 25 y 34 años). Por otro lado, los grupos de edad entre 45 y 54 sufren ligeros aumentos y mientras la población de más edad (a partir de los 55 años) ha aumentado significativamente. Esto se debe a que la población de la CAV está envejeciendo, con tasas de natalidad relativamente bajas.

Esta situación genera que el número de personas mayores cada vez supere en mayor medida al de los jóvenes. Por ejemplo, en 2005 el grupo de edad mayoritario era el comprendido entre 30 y 34 años, mientras que en 2020 dicho grupo de edad es casi el menos común (solo superado por el de 25-29 años), y el grupo de edad con mayor población es 45-49 años, grupo que antes ocupaba el cuarto lugar.

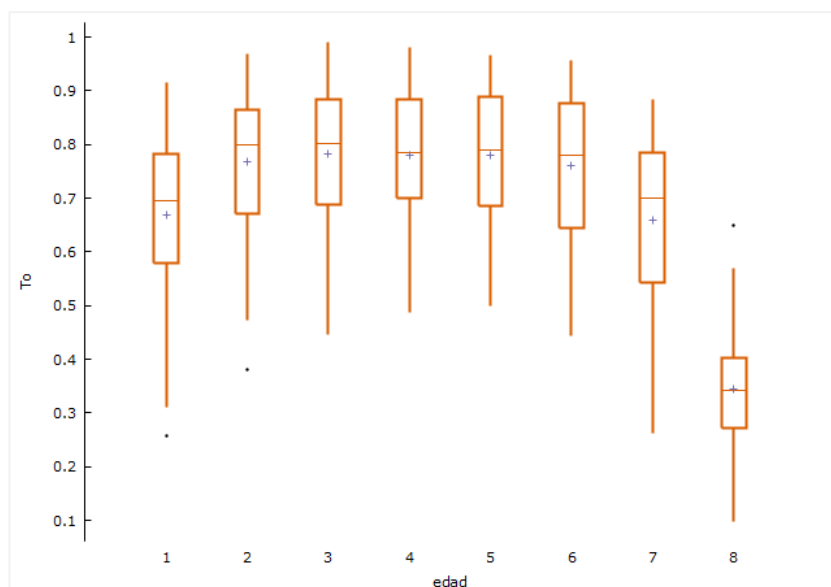
Figura 4: Pirámide poblacional de los años 1991 y 2021



Fuente: Eustat: censos de población y viviendas

Aunque los datos reflejados en esta pirámide poblacional sean anteriores al 2005 (objeto de este trabajo), y recojan datos a nivel estatal, se puede observar el mismo fenómeno de la figura 3: una población cada vez más envejecida.

Figura 5: Distribución de tasa de ocupación por edad



Fuente: Elaboración propia a partir de los datos de la EPA (INE)

Al analizar la relación entre la edad y la tasa de ocupación, se observa que conforme la edad aumenta, aumenta la tasa de ocupación, hasta que a partir de los 50 años la tasa comienza a descender (ver figura 5).

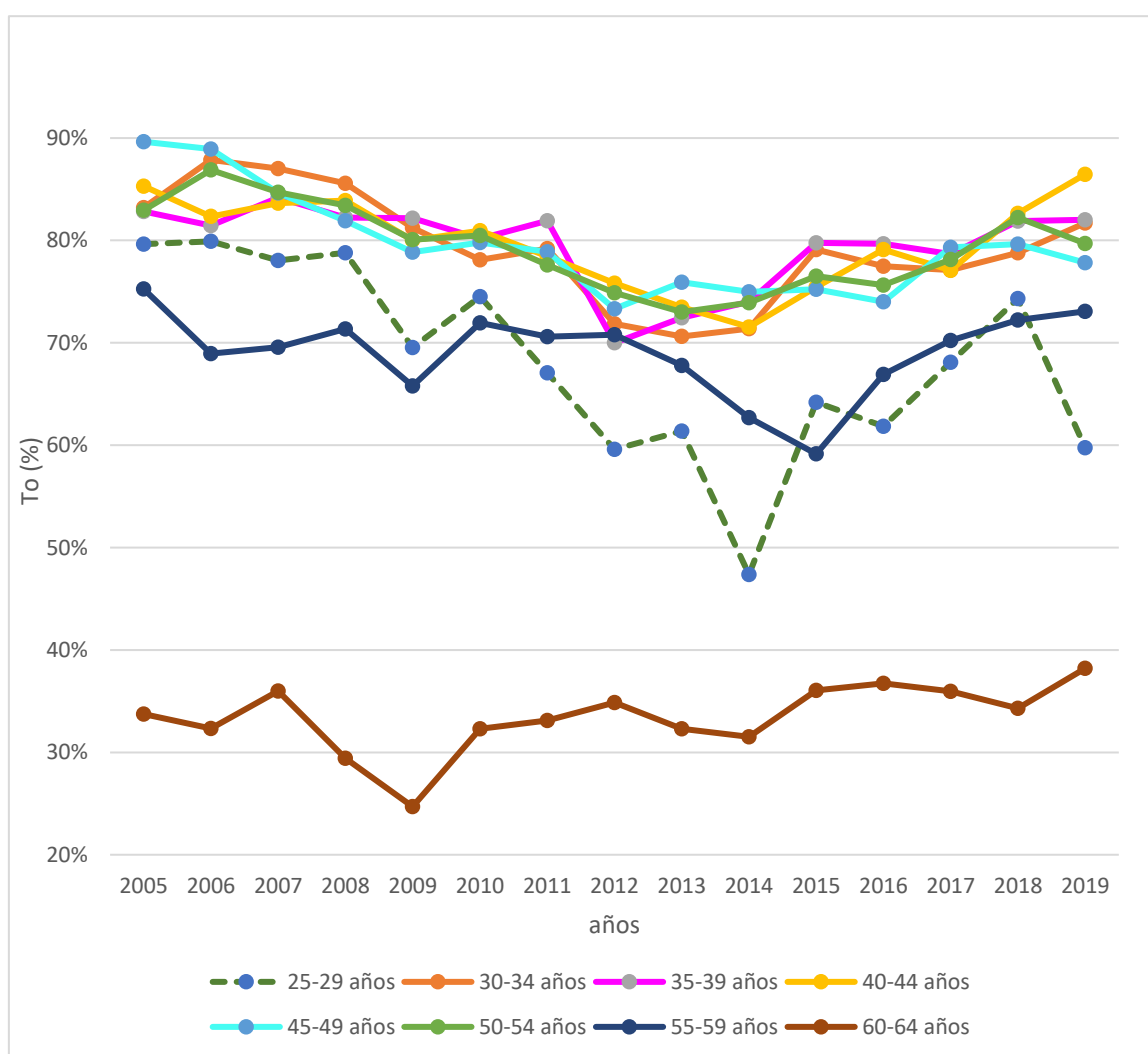
Por un lado, esto es consecuencia de que los individuos más jóvenes retrasan su incorporación al mercado laboral debido a que invierten sus primeros años en su formación (si analizáramos las tasas en menores de 25 años, esto sería mucho más notable). Además, la búsqueda del primer empleo puede ser un proceso complejo debido a que para acceder a muchos de los empleos que se ofertan en el mercado laboral se exige cierta experiencia, experiencia de la cual los más jóvenes carecen.

Por otro lado, la tasa de ocupación sufre una bajada a partir del grupo 7 (entre 55 y 59 años), con unas tasas menores del 35% en el grupo 8 (entre 60 y 64 años). Aunque en España la edad de jubilación sea de 66 años, existe la posibilidad de prejubilarse a los 61 (cumpliendo ciertas condiciones), e incluso en ciertos casos quienes estén en un periodo largo de desempleo y hayan agotado su prestación contributiva, también podrán optar a un subsidio para mayores de 52 años hasta la edad de jubilación, quedando de esta manera también fuera de la población ocupada. Por ello no es de extrañar que la tasa de ocupación a partir del grupo 6 (entre 50 y 54 años) comiencen a decaer y en los siguientes dos grupos de edad sufra una disminución mucho mayor, de hecho, las tasas de ocupación se reducen a prácticamente la mitad. Por ejemplo, en el caso de las mujeres con estudios de orientación profesional, en 2017 su tasa de ocupación era de un 75,59% para el grupo comprendido entre los 50 y 54 años, bajando dicha tasa para el grupo de mujeres de 60 y 64 años hasta el 31,85%.

Se deberá tener en cuenta que a partir de los 45-50 años las tasas de ocupación van cayendo a medida que los individuos van cumpliendo años, porque van retirándose poco a poco del mercado laboral, ya sea porque debido a causas físicas no pueden realizar una jornada laboral, o simplemente porque pueden permitírselo. Además, quienes pierden su empleo a una edad tardía, suelen tener dificultades a la hora de encontrar un nuevo empleo y en muchos casos podrán tener complicado el adaptarse a las nuevas metodoloías de trabajo, por lo que quedarán también en muchos casos fuera del mundo laboral.

De esta manera, el grupo de edad que no solo representa una mayor parte de la población, sino que además recibe tasas de ocupación mas elevadas es la franja de edad entre los 40 y 49 años de edad.

Figura 6: Tasa de ocupación por edades a lo largo del tiempo



Fuente: Elaboración propia a partir de los datos de la EPA (INE)

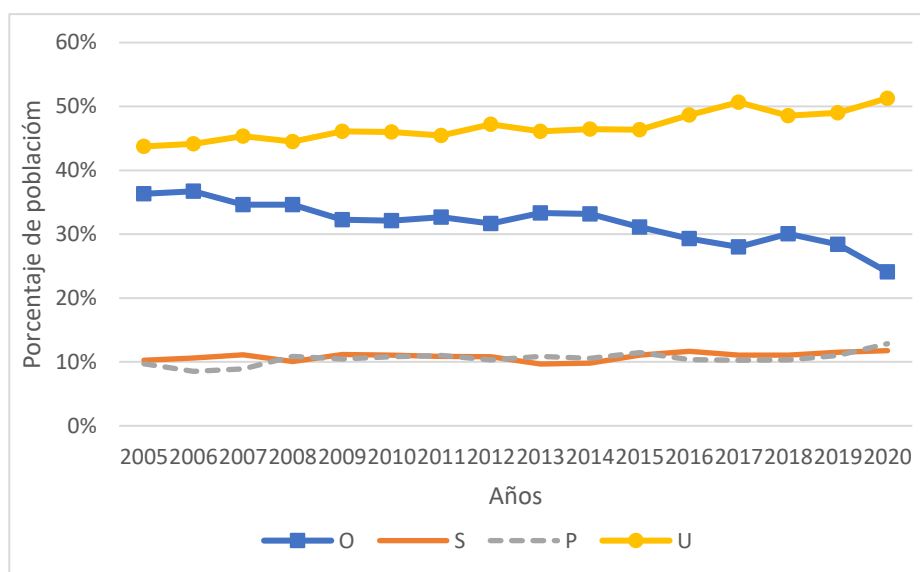
Existen ciertas tendencias que se mantienen año tras año, como por ejemplo que el grupo con mayor edad siempre es el que registra tasas de ocupación más bajas, seguido por justo el intervalo de edad anterior (55-59 años) y el intervalo más joven (25-29 años). Como se puede comprobar en la figura 6, las tasas de ocupación de los grupos de edades intermedias (desde los 30 a 54 años) son las que tienen tasas de ocupación mayores que las del resto de edades. Pese a que en este análisis los grupos de edad entre 30 y 54 años se clasifiquen en cinco grupos quinquenales, tienen tendencias similares y tasas de ocupación similares, de hecho, más adelante en este análisis se agruparán en únicamente cuatro categorías.

Cabe destacar, como ya se ha comentado en otros apartados, que el empleo juvenil es el que en mayor medida sufre los cambios del mercado laboral, debido a la elevada precariedad e inestabilidad. Por ello sufre grandes altibajos a lo largo del tiempo, llegando a tener una diferencia de 32,54 puntos porcentuales entre su tasa de ocupación máxima y mínima.

Por otro lado, el grupo de edad de 50 a 54 años es el que más estable se mantiene a lo largo del tiempo, y es más constante ante las variaciones del mercado laboral, aunque no por ello sea el que siempre presente las tasas de ocupación más altas.

3.1.3 Nivel de formación

Figura 7: Porcentaje de individuos por cada nivel de estudios a lo largo de los años



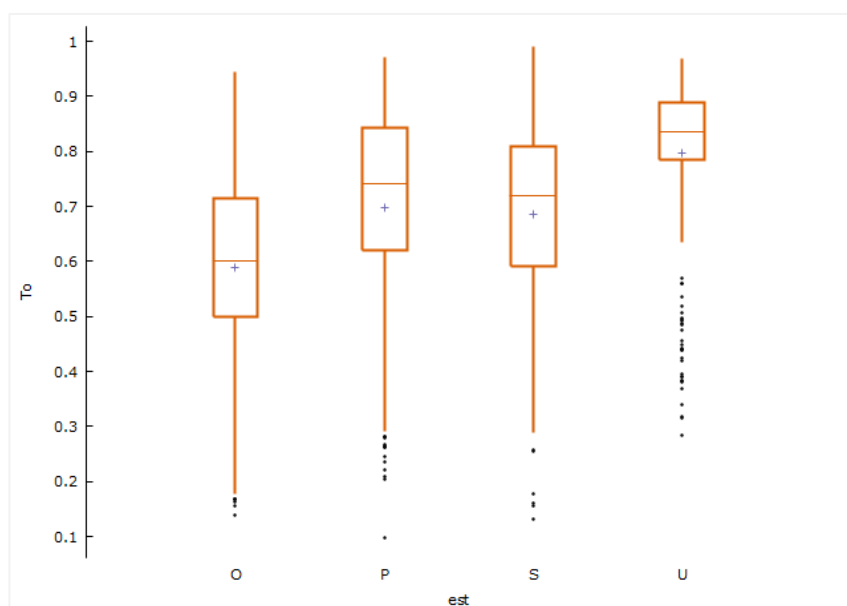
Fuente: Elaboración propia a partir de los datos de la EPA (INE)

Esta figura nos muestra como está distribuido el nivel de estudios en la población, siendo llamativo que el porcentaje de población que representa a los individuos con nivel de estudios de orientación profesional y orientación general se haya mantenido estable con valores en torno al 10%. Y, por otro lado, mientras que la educación está sufriendo ligeros descensos, es la

educación universitaria la que representa a un mayor porcentaje de población, y además, con tendencia al alza. Esta tendencia se debe a la situación del mercado laboral, y es que las altas tasas de ocupación entre los individuos con formación superior, puede ser un incentivo para que la población opte por este tipo de educación.

En 2005 en la CAV había alrededor de 496.611 individuos cuyos niveles máximos de educación eran de formación obligatoria (un 39,7% de la población), mientras que en 2020 el número se había reducido a unos 280.318 individuos (el 23,9% de la población). Por otro lado, el nivel de estudios mayoritario entre la población activa en la CAV a lo largo de todos los años analizados es el nivel de educación superior.

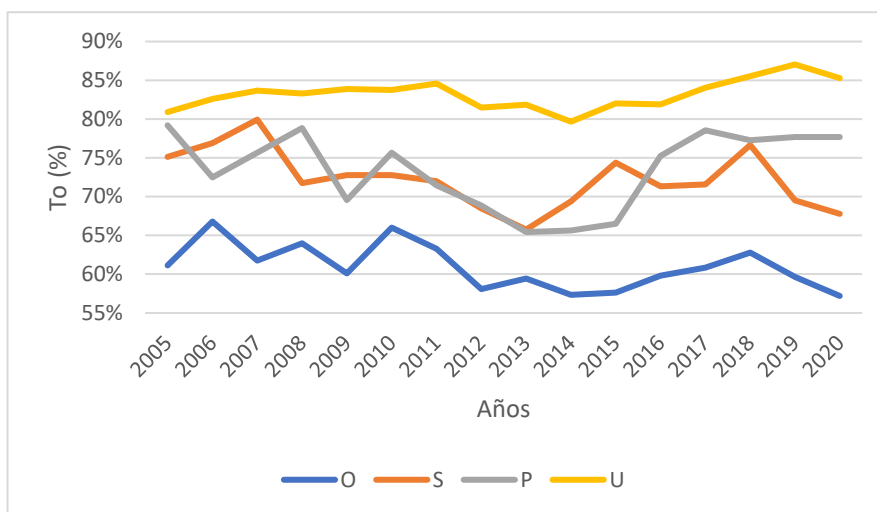
Figura 8: Distribución de la tasa de ocupación por nivel de estudios



Fuente: Elaboración propia a partir de los datos de la EPA (INE)

A la hora de analizar la tasa de ocupación respecto al nivel de estudios, podemos afirmar que los individuos con estudios universitarios son los que mayores tasas de ocupación presentan de manera general, y en cambio, quienes tienen únicamente estudios obligatorios son los que presentan tasas menores de ocupación, además de una mayor dispersión. Por otro lado, los individuos de nivel de formación de orientación profesional y los de orientación general tienen tasas de ocupación bastante similares.

Figura 9: Tasa de ocupación para cada nivel de formación



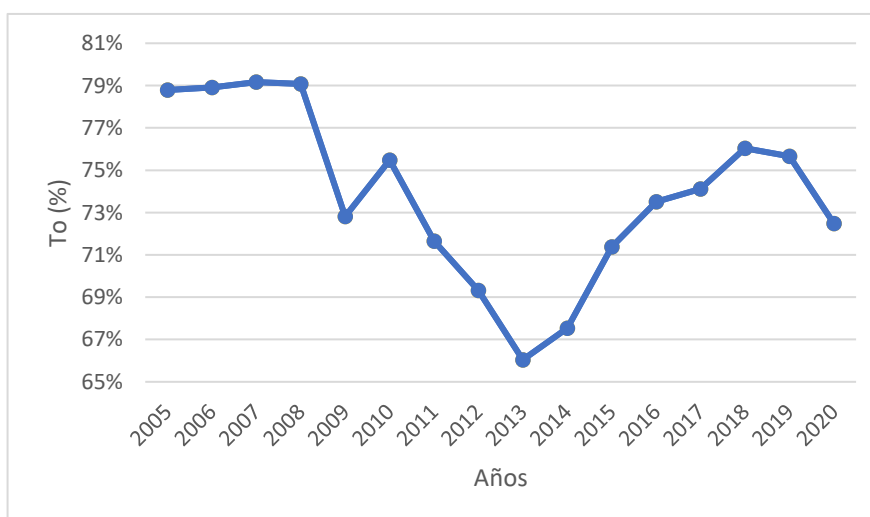
Fuente: Elaboración propia a partir de los datos de la EPA (INE)

Al observar la evolución de la tasa de ocupación para cada nivel de formación (figura 9) podemos confirmar lo observado en la figura 8, que de manera general los individuos con formación obligatoria tendrán tasas de ocupación menores, y los que tengan formación superior tendrán tasas más elevadas, además de más constantes a lo largo del tiempo.

3.2 Variable endógena

3.2.1 Tasa de ocupación

Figura 10: Evolución temporal de la tasa de ocupación (2005-2020)



Fuente: Elaboración propia a partir de los datos de la EPA (INE)

En la figura 10 se presenta la evolución temporal de la tasa de ocupación entre los años 2005 y 2020. Se deberá tener en cuenta que en la figura 10 aparecen reflejadas las tasas de ocupación medias, pero como se ha comentado en los apartados anteriores, las tasas de ocupación varían según la población.

La tasa de ocupación presenta cambios continuos a lo largo del tiempo, con un valor medio del 69,25%, logrando su valor medio máximo en 2008 con una tasa de ocupación media del 79%, tasa que se vio disminuida tras la crisis financiera del 2008. Por otro lado, debido a la mencionada crisis del 2008, en 2013 la tasa de empleo obtuvo su valor medio mínimo, con una tasa del 66%. Podemos observar también como la tasa de ocupación sufre una bajada de 2019 a 2020 debido a los efectos del COVID-19.

Tabla 3: Estadísticos principales

Estadísticos principales de la tasa de ocupación

<i>Media</i>	0.69250
<i>Mediana</i>	0.73538
<i>Mínimo</i>	0.098039
<i>Máximo</i>	0.99130
<i>Desviación típica</i>	0.18941
<i>Percentil del 5%</i>	0.31565
<i>Percentil del 95%</i>	0.93391
<i>Rango intercuartílico</i>	0.25319

Fuente: Elaboración propia a partir de los datos de la EPA (INE)

Respecto a los estadísticos principales reflejados en la tabla 3, cabe destacar que el percentil del 5% tiene una tasa del 31,56%, por lo que solamente el 5% de los pseudo-individuos tendrán tasas menores a esa. La mitad de los pseudo-individuos tendrán tasas de ocupación mayores al 73,53% y tan solo el 5% de los pseudo-individuos tendrán tasas mayores al 93,39%.

4. METODOLOGÍA

4.1 Modelo especificado

A través de las variables explicativas anteriormente mencionadas (sexo, edad y nivel de estudios) se intentará explicar la variable endógena: la tasa de ocupación. De este modo se estimará el efecto de la incidencia de dichas variables en la tasa de ocupación de la CAV.

Para ello, se especificará un modelo de regresión lineal, y al tratarse de variables de carácter cualitativo, se deberán crear variables ficticias para cada una de ellas. Las variables ficticias sirven para incorporar al modelo econométrico variables cualitativas (Alonso et al., 2005). De esta manera se especifica un modelo econométrico que incluye las siguientes variables:

4.1.1 Variable endógena:

- To_{it} : La tasa de ocupación del individuo i en el año t .

4.1.2 Variables exógenas:

- Sexo del individuo
 - V_{it} : Si el individuo es hombre =1, en caso contrario 0.
 - M_{it} : Si la individuo es mujer =1, en caso contrario 0.
- Edad del individuo
 - $E1_{it}$: Si el individuo tiene una edad comprendida entre los 25 y 29 años = 1, en caso contrario = 0
 - $E2_{it}$: Si el individuo tiene una edad comprendida entre los 30 y 34 años = 1, en caso contrario = 0
 - $E3_{it}$: Si el individuo tiene una edad comprendida entre los 35 y 39 años = 1, en caso contrario = 0
 - $E4_{it}$: Si el individuo tiene una edad comprendida entre los 40 y 44 años = 1, en caso contrario = 0
 - $E5_{it}$: Si el individuo tiene una edad comprendida entre los 45 y 49 años = 1, en caso contrario = 0
 - $E6_{it}$: Si el individuo tiene una edad comprendida entre los 50 y 54 años = 1, en caso contrario = 0
 - $E7_{it}$: Si el individuo tiene una edad comprendida entre los 55 y 59 años = 1, en caso contrario = 0
 - $E8_{it}$: Si el individuo tiene una edad comprendida entre los 60 y 64 años = 1, en caso contrario = 0
- Nivel de formación:
 - O_{it} : Si el individuo tiene un nivel de formación de estudios obligatorios = 1, en caso contrario = 0

- S_{it} : Si el individuo tiene un nivel de formación de orientación general = 1, en caso contrario = 0
- P_{it} : Si el individuo tiene un nivel de formación de orientación profesional = 1, en caso contrario = 0
- U_{it} : Si el individuo tiene un nivel de formación de educación superior = 1, en caso contrario = 0
- Año al que pertenece la tasa de ocupación
 - $A05_{it}$: Si los datos pertenecen al año 2005, en caso contrario = 0
 - $A06_{it}$: Si los datos pertenecen al año 2006, en caso contrario = 0
 - $A07_{it}$: Si los datos pertenecen al año 2007, en caso contrario = 0
 - $A08_{it}$: Si los datos pertenecen al año 2008, en caso contrario = 0
 - $A09_{it}$: Si los datos pertenecen al año 2009, en caso contrario = 0
 - $A10_{it}$: Si los datos pertenecen al año 2010, en caso contrario = 0
 - $A11_{it}$: Si los datos pertenecen al año 2011, en caso contrario = 0
 - $A12_{it}$: Si los datos pertenecen al año 2012, en caso contrario = 0
 - $A13_{it}$: Si los datos pertenecen al año 2013, en caso contrario = 0
 - $A14_{it}$: Si los datos pertenecen al año 2014, en caso contrario = 0
 - $A15_{it}$: Si los datos pertenecen al año 2015, en caso contrario = 0
 - $A16_{it}$: Si los datos pertenecen al año 2016, en caso contrario = 0
 - $A17_{it}$: Si los datos pertenecen al año 2017, en caso contrario = 0
 - $A18_{it}$: Si los datos pertenecen al año 2018, en caso contrario = 0
 - $A19_{it}$: Si los datos pertenecen al año 2019, en caso contrario = 0
 - $A20_{it}$: Si los datos pertenecen al año 2020, en caso contrario = 0

Se deberá tener en cuenta que en caso de que una variable explicativa sea una combinación lineal de alguna de las otras variables explicativas, dará lugar a un problema de multicolinealidad. Esto ocurre cuando se trabaja con variables ficticias. Por ello, se excluirá una variable ficticia de cada variable cualitativa para evitar este tipo de problemas. En este caso quedarán excluidas del modelo econométrico y por lo tanto servirán como referencia las siguientes variables:

- Sexo: (V_{it}) Si el individuo es mujer.
- Edad: (EA_{it}) Si la edad del individuo está entre 40 y 44 años.
- Nivel de formación: (O_{it}) Si el nivel de formación del individuo es de educación obligatoria.
- Año: ($A19_{it}$) Si los datos pertenecen al año 2019.

De este modo, se especificará el siguiente modelo econométrico:

Modelo inicial [0]

$$\begin{aligned}
 T_{oit} = & \beta_1 + \beta_2 V_{it} + \beta_3 S_{it} + \beta_4 P_{it} + \beta_5 U_{it} + \beta_6 E1_{it} + \beta_7 E2_{it} + \beta_8 E3_{it} + \beta_9 E5_{it} \\
 & + \beta_{10} E6_{it} + \beta_{11} E7_{it} + \beta_{12} E8_{it} + \beta_{13} A05_{it} + \beta_{14} A06_{it} + \beta_{15} A07_{it} \\
 & + \beta_{16} A08_{it} + \beta_{17} A09_{it} + \beta_{18} A10_{it} + \beta_{19} A11_{it} + \beta_{20} A12_{it} + \beta_{21} A13_{it} \\
 & + \beta_{22} A14_{it} + \beta_{23} A15_{it} + \beta_{24} A16_{it} + \beta_{25} A17_{it} + \beta_{26} A18_{it} + \beta_{27} A20_{it} + U_{it}
 \end{aligned}$$

$$t = 1, 2, \dots, 16$$

$$i = 1, 2, \dots, 64$$

4.2 Hipótesis

Este modelo tendrá como objetivo cuantificar la relación existente entre la variable Tasa de Ocupación y las variables explicativas sexo, edad, formación y tiempo. Se supone que este modelo cumple las hipótesis básicas habituales de los modelos de regresión:

- **H1-** Hipótesis sobre la relación: el modelo está bien especificado, todas las variables explicativas del modelo sirven para explicar la variable endógena, y ninguna variable que explique la variable a explicar queda fuera del modelo.
- **H2-** Hipótesis respecto a los coeficientes: los parámetros son constantes en toda la muestra y aparecen de forma lineal.
- **H3-** Hipótesis sobre la perturbación aleatoria:
 1. Media cero: $E(U_i) = 0$
 2. Homocedasticidad: $Var(U_i) = \sigma^2 \quad \{ \forall_i \}$
 3. No autocorrelación: $Cov(U_i, U_j) = 0, \quad \{ \forall_{i \neq j} \}$
 4. Normalidad: $U_i \sim N(0, \sigma^2)$

Hipótesis de las variables explicativas:

- Pueden considerarse fijas en sucesivas muestras, son variables no aleatorias.
- Las variables son cuantitativas.
- Las variables explicativas son linealmente independientes entre sí, es decir, no son combinaciones lineales exactas ni pueden expresarse como combinación lineal del resto.

Dadas estas hipótesis básicas el estimador de mínimos cuadrados ordinarios (MCO) de los coeficientes del modelo $\hat{\beta}_{MCO} = (X'X)^{-1}X'Y$ tendrá las siguientes propiedades:

- Linealidad: el estimador es una combinación lineal de las perturbaciones.
- Insesgaredad: $E(\hat{\beta}) = \beta$, por lo que $\hat{\beta}$ es insesgado.

- Eficiencia: si se cumplen las hipótesis básicas, el teorema de Gauss-Markov establece que el estimador de MCO de los coeficientes del modelo es el estimador lineal e insesgado de menor varianza y, en ese sentido es el óptimo.

$$\hat{\beta} = ELIO = E_{\text{estimador}} L_{\text{ineal}} I_{\text{insesgado}} O_{\text{ptimo}}$$

- Matriz de covarianzas: la matriz de varianzas y covarianzas es $Var(\beta_{MCO}) = \sigma^2(X'X)^{-1}$

Dado como se han construido los datos dando lugar a un pseudo-panel, los datos no van a ser independientes en el tiempo y tendrán una heterocedastidad de cohorte específica, con lo que no cumplen las hipótesis básicas de homocedasticidad y no correlación.

Los datos empleados en este modelo econométrico pertenecen a pseudo-individuos muy diversos entre sí y por ello $Var(U_{it}) \neq \sigma^2$, ya que la dispersión que tendrán ciertos pseudo-individuos será mucho mayor que la que tengan otros, es decir, en este caso no hay homocedasticidad.

Por otro lado, es probable la existencia de autocorrelación, ya que al tratarse de un panel de datos se realiza un seguimiento de una variable a lo largo del tiempo. De esta forma, los resultados obtenidos en el año t estarán relacionados con los del año $t - 1$, lo que ocurre también en las perturbaciones. Por ello, $Cov(U_{it}, U_{is}) \neq 0$.

Bajo estas circunstancias, el estimador de MCO pierde algunas de sus propiedades, mientras que sigue siendo lineal e insesgado, ya no será eficiente ya que no será el de menor varianza. Además, se estimarán de forma incorrecta las varianzas del estimador de MCO, aunque las estimaciones de las β sean insesgadas. Para casos como este dónde hay heterosticidad y autocorrelación con datos de panel se empleará el estimador que propone Arellano (1987) para estimar la matriz de varianzas y covarianzas del estimador de MCO. El estimador de la matriz de de varianzas y covarianzas del estimador MCO que propone Arellano sólo tiene propiedades asintóticas, en muestras grandes, por lo que sólo se debe aplicar cuando se cuenta con gran cantidad de datos, como es en este caso.

Este modelo se estimará entonces mediante MCO y las estimaciones de las desviaciones típicas se calcularán a través de las correcciones que propone Arellano. Para ello se empleará el software econométrico GRETL aplicando las desviaciones típicas robustas (HAC).

4.3 Contrastes de hipótesis

Una vez especificado el modelo inicial, se podrán/deberán realizar tanto contrastes simples como múltiples respecto a los parámetros. Como ya se ha mencionado en el apartado anterior, las estimaciones realizadas en este TFG serán realizados con el estimador de la matriz de varianzas y covarianzas del estimador MCO propuesto por Arellano, por lo que los contrastes que se describen a continuación serán contrastes asintóticos.

4.3.1 Hipótesis simples

Las hipótesis simples son aquellas que se formulan a través de una única igualdad. Para llevar a cabo un contraste de significación simple, antes de nada, se deberán establecer una hipótesis nula y una alternativa y construir un estadístico para contrastar las hipótesis formuladas. En este caso se emplearán las siguientes hipótesis y estadísticos:

$$\begin{cases} H_0: \beta_i = 0 \\ H_a: \beta_i \neq 0 \end{cases} \quad t = \frac{\hat{\beta} - 0}{\hat{\sigma}_{\hat{\beta}}^{\text{Arellano}}} \xrightarrow{H_0} N(0,1)$$

En caso de cumplirse la hipótesis nula (H_0) la variable correspondiente no sería significativa, lo que indicaría que los cambios de dicha variable no influirían a la variable endógena. En caso contrario, si se cumple hipótesis nula (H_0) la variable en cuestión si sería significativa y afectaría a la variable endógena.

Regla de decisión

- Si $|t| > t_{\alpha/2}$ se rechaza H_0 .
- Si $|t| \leq t_{\alpha/2}$ no se rechaza H_0 .

Si $|t| \leq t_{0,05/2} = 1,96$; se acepta H_0 frente a H_a al 5% de significación. En este caso la variable no sería relevante. En caso contrario, si $|t| > t_{\alpha/2}$, dicha variable si será significativa e influirá en la variable endógena.

4.3.2 Hipótesis múltiples

Por otro lado, las hipótesis que contrastan más de una restricción se considerarán hipótesis múltiples. En ocasiones conviene llevar a cabo también contrastes múltiples para comprobar si un conjunto de variables independientes tiene efecto sobre la variable dependiente. El estimador que se empleará será el siguiente:

$$\begin{cases} H_0: R\mathbb{B} = r \\ H_a: R\mathbb{B} \neq r \text{ no es cierta} \end{cases} \quad F_c = \frac{(SCR_R - SCR_{NR})}{SCR_{NR}/(N - K)} = F_c \xrightarrow{H_0} \chi_{q|}^2$$

Donde:

- SCR_R = Suma de cuadrados residuales del modelo restringido (modelo en el que se incorporan las restricciones $R\beta$).
- SCR_{RN} = Suma de cuadrados residuales del modelo no restringido.
- q = número de variables a contrastar.
- $N - K$ = grado de libertad.



Se diferenciarán dos modelos diferentes, el modelo no restringido (NR) será el modelo inicial y el no restringido será el modelo en el que H_0 se cumpla. En este tipo de modelos el modelo restringido siempre tendrá un menor número de parámetros que el modelo no restringido.

Regla de decisión

- Si $F > \chi_{q|1\alpha}^2$ se rechaza H_0
- Si $F \leq \chi_{q|1\alpha}^2$ no se rechaza H_0

Para un nivel de significación α , en caso de que $F > \chi_{q|1\alpha}^2$ se rechazará la hipótesis nula, por lo que las variables correspondientes serán conjuntamente significativas para dicho nivel de significación. En caso contrario, $F \leq \chi_{q|1\alpha}^2$, no se rechazará, con lo que la variable no será conjuntamente significativa.

5. MODELOS ESTIMADOS

5.1 Modelo para los años 2005-2020

5.1.1 Estimaciones del modelo inicial

Modelo inicial [0]

$$\begin{aligned}
 T_{oit} = & \beta_1 + \beta_2 V_{it} + \beta_3 S_{it} + \beta_4 P_{it} + \beta_5 U_{it} + \beta_6 E1_{it} + \beta_7 E2_{it} + \beta_8 E3_{it} + \beta_9 E5_{it} \\
 & + \beta_{10} E6_{it} + \beta_{11} E7_{it} + \beta_{12} E8_{it} + \beta_{13} A05_{it} + \beta_{14} A06_{it} + \beta_{15} A07_{it} \\
 & + \beta_{16} A08_{it} + \beta_{17} A09_{it} + \beta_{18} A10_{it} + \beta_{19} A11_{it} + \beta_{20} A12_{it} + \beta_{21} A13_{it} \\
 & + \beta_{22} A14_{it} + \beta_{23} A15_{it} + \beta_{24} A16_{it} + \beta_{25} A17_{it} + \beta_{26} A18_{it} + \beta_{27} A20_{it} + U_{it}
 \end{aligned}$$

$t = 1, 2, \dots, 16$
 $i = 1, 2, \dots, 64$

El modelo [0] propuesto en el modelo anterior se ha estimado por MCO con la corrección de Arellano a la hora de estimar las desviaciones típicas obteniendo así los siguientes resultados:

Tabla 4: Resultado de las estimaciones del modelo inicial

Variable	Coficiente	Desv. típica	Estadístico t	Valor p	
const	0,633984	0,0134271	47,22	6,86e-051	***
V	0,130450	0,0100284	13,01	1,66e-019	***
S	0,108910	0,0156780	6,947	2,47e-09	***
P	0,0958507	0,0124180	7,719	1,11e-010	***
U	0,208747	0,0126928	16,45	1,77e-024	***
E1	-0,111387	0,0249320	-4,468	3,36e-05	***
E2	-0,0106481	0,0147130	-0,7237	0,4719	
E3	0,00164144	0,0108265	0,1516	0,8800	
E5	-0,000203436	0,0104535	-0,01946	0,9845	
E6	-0,0194501	0,0157341	-1,236	0,2210	
E7	-0,120922	0,0179520	-6,736	5,76e-09	***
E8	-0,436237	0,0168420	-25,90	2,70e-035	***
A05	-0,00256878	0,0158328	-0,1622	0,8716	
A06	0,00730668	0,0159591	0,4578	0,6486	
A07	0,00793277	0,0146587	0,5412	0,5903	
A08	-0,00342372	0,0145816	-0,2348	0,8151	
A09	-0,0385449	0,0121163	-3,181	0,0023	***
A10	-0,0199874	0,0115596	-1,729	0,0887	*

A11	-0,0212301	0,0105602	-2,010	0,0487	**
A12	-0,0570927	0,0102107	-5,591	5,19e-07	***
A13	-0,0709866	0,0105521	-6,727	5,96e-09	***
A14	-0,0625865	0,00891598	-7,020	1,85e-09	***
A15	-0,0460395	0,0097835	-4,728	1,32e-05	***
A16	-0,0257934	0,00955246	-2,700	0,0089	***
A17	-0,0145245	0,0110373	-1,316	0,1930	
A18	-0,00215421	0,00873854	-0,2465	0,8061	
A20	-0,0173356	0,0105779	-1,639	0,1062	
<i>R-cuadrado corregido</i>			0,831482		

Fuente: Elaboración propia a partir de los datos de la EPA (INE)

A continuación, se realizan los contrastes de significatividad individual y conjunta para determinar que variables resultarán significativas en el cálculo de la tasa de ocupación.

Para empezar, se llevarán a cabo los contrastes respectivos a la edad:

$$\begin{cases} H_0: \beta_6 = 0 \\ H_a: \beta_6 \neq 0 \end{cases} \quad t = \frac{\hat{\beta} - 0}{\hat{\sigma}_{\hat{\beta}}^{\text{Arellano}}} \xrightarrow{H_0} N(0,1)$$

Al calcular el estadístico t, obtenemos que $|t| = 4,468$, por lo tanto H_0 se rechaza a un nivel de significación del %5, ya que $|t| = 4,468 > t_{0,05/2} = 1,96$. Como consecuencia, la variable E1 es significativa a la hora de calcular las tasas de ocupación.

$$\begin{cases} H_0: \beta_7 = 0 \\ H_a: \beta_7 \neq 0 \end{cases} \quad t = \frac{\hat{\beta} - 0}{\hat{\sigma}_{\hat{\beta}}^{\text{Arellano}}} \xrightarrow{H_0} N(0,1)$$

En este caso, obtenemos un valor de $|t| = 0,7237$ para el estadístico t, por lo tanto H_0 se acepta a un nivel de significación del %5, ya que $|t| = 0,7237 \leq t_{0,05/2} = 1,96$. Como consecuencia, la variable E2 es relevante.

Siguiendo el mismo proceso con todas y cada una de las variables ficticias correspondientes a la edad, obtenemos que mientras que las variables E1, E7 y E8 son relevantes, las variables E2, E3, E5 y E6 no son relevantes para analizar la tasa de ocupación, debido a que los valores absolutos del “estadístico t” son inferiores a $t_{\alpha/2}$. Es por ello por lo que, a partir de este punto, hasta el final de este TFG, no se incluirán dichas variables. De esta forma se diferenciarán cuatro grupos de edad diferentes: 25-29 años, 30- 54 años, 55-59 años y 60-64 años, siendo el segundo tramo de edad la variable de referencia. Llevando a cabo el contraste de significatividad del conjunto de estas variables obtenemos unos resultados similares:

$$\begin{cases} H_o: \beta_7 = \beta_8 = \beta_9 = \beta_{10} = 0 \\ H_a: \beta_7 \neq \beta_8 \neq \beta_9 \neq \beta_{10} \neq 0 \end{cases} F_c = \frac{SCR_R - SCR_{NR}}{SCR_{NR}/N - K} \xrightarrow{H_o} \chi^2_{q1}$$

En este contraste múltiple, el estadístico toma un valor de $F_c = 3,7945$, y al tener un valor inferior al valor de $\chi^2_{4|0,05} = 9,48$, no se rechaza H_o , por lo que no son relevantes dichas variables conjuntamente. De esta forma obtenemos los mismos resultados obtenidos a través del contraste simple.

Respecto a las variables correspondientes a cada año, es cierto que algunas no son significativas, ya que la diferencia que pueden tener con el año de referencia es pequeña, pero para analizar de forma efectiva la evolución temporal se ha decidido no eliminarlas.

El resto de variables, las correspondientes a las variables del sexo y nivel de estudios, se mantendrán en este análisis, debido a que como se puede observar en la columna del estadístico t de la tabla 4, todas las variables tendrán un estadístico $|t| > t_{0,05/2} = 1,96$, por lo que son variables relevantes.

5.1.1 Modelo inicial para las variables relevantes

A continuación se han eliminado las variables no significativas del modelo [0], obteniendo así el modelo [1]:

Modelo [1]

$$\begin{aligned} To_{it} = & \beta_1 + \beta_2 V_{it} + \beta_3 S_{it} + \beta_4 P_{it} + \beta_5 U_{it} + \beta_6 E1_{it} + \beta_7 E7_{it} + \beta_8 E8_{it} + \beta_9 A05_{it} \\ & + \beta_{10} A06_{it} + \beta_{11} A07_{it} + \beta_{12} A08_{it} + \beta_{13} A09_{it} + \beta_{14} A10_{it} + \beta_{15} A11_{it} \\ & + \beta_{16} A12_{it} + \beta_{17} A13_{it} + \beta_{18} A14_{it} + \beta_{19} A15_{it} + \beta_{20} A16_{it} + \beta_{21} A17_{it} \\ & + \beta_{22} A18_{it} + \beta_{23} A20_{it} + U_{it} \end{aligned}$$

$$t = 1, 2, \dots, 16$$

$$i = 1, 2, \dots, 32$$

Tabla 5: Resultado de las estimaciones del modelo inicial para las variables significativas para el periodo 2005-2020

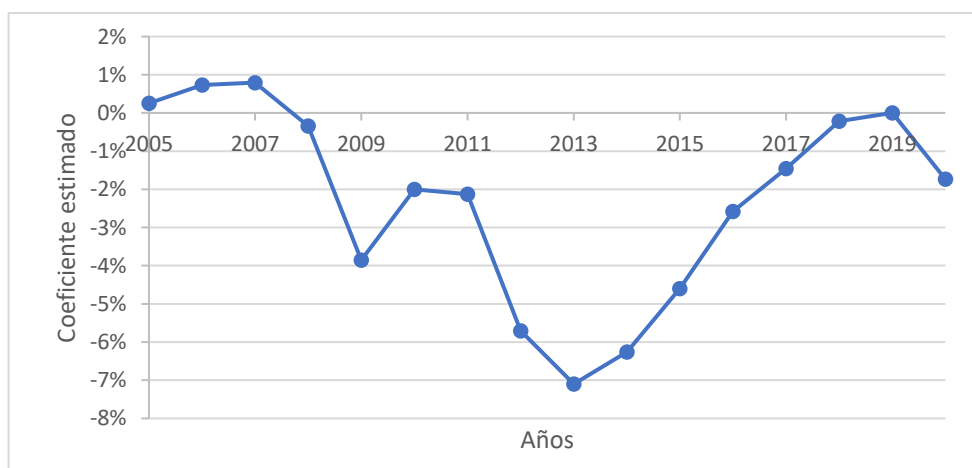
Variable	Coficiente	Desv. típica	Estadístico t	Valor p	
const	0,628252	0,0126522	49,66	<0,0001	***
V	0,130450	0,0101412	12,86	<0,0001	***
S	0,108910	0,0159057	6,847	<0,0001	***
P	0,0958507	0,0127175	7,537	<0,0001	***
U	0,208747	0,0129871	16,07	<0,0001	***

E1	-0,105655	0,0243085	-4,346	<0,0001	***
E7	-0,115190	0,0171105	-6,732	<0,0001	***
E8	-0,430504	0,0159468	-27,00	<0,0001	***
A05	-0,00256878	0,0158012	-0,1626	0,8714	
A06	0,00730668	0,0159272	0,4588	0,6480	
A07	0,00793277	0,0146294	0,5422	0,5896	
A08	-0,00342372	0,0145524	-0,2353	0,8148	
A09	-0,0385449	0,0120921	-3,188	0,0022	***
A10	-0,0199874	0,0115365	-1,733	0,0881	*
A11	-0,0212301	0,0105391	-2,014	0,0482	**
A12	-0,0570927	0,0101902	-5,603	<0,0001	***
A13	-0,0709866	0,0105310	-6,741	<0,0001	***
A14	-0,0625865	0,00889815	-7,034	<0,0001	***
A15	-0,0460395	0,00971887	-4,737	<0,0001	***
A16	-0,0257934	0,00953336	-2,706	0,0088	***
A17	-0,0145245	0,0110153	-1,319	0,1921	
A18	-0,00215421	0,00872107	-0,2470	0,8057	
A20	-0,0173356	0,0105568	-1,642	0,1055	
<i>R-cuadrado</i>				0,834613	

Fuente: Elaboración propia a partir de los datos de la EPA (INE)

Esta especificación del modelo permite estimar la evolución temporal, en base al año 2019, de la tasa de ocupación en el periodo analizado.

Figura 11: Estimación temporal de la tasa de ocupación en el modelo inicial



Fuente: Elaboración propia a partir de los datos de la EPA (INE)

La evolución estimada es similar a la evolución temporal de la tasa de ocupación reflejada en la figura 10 (ver página 27).

Como se puede observar en la figura 11, el efecto del COVID-19 no se limita únicamente a la caída del 1,7% que supuso el 2020, sino que la inercia de crecimiento que se llevaba observando desde 2013 se detuvo, por lo que el efecto de la pandemia en la tasa de ocupación fue mayor del 1,7% reflejado en el modelo inicial.

Para poder analizar correctamente la verdadera influencia del COVID-19, se deberá modelizar la evolución temporal, proponiendo una especificación del modelo que permita estimar la evolución temporal de la tasa de ocupación. En los siguientes apartados se propondrán distintos modelos alternativos con intención de modelizar de distintas formas la evolución temporal de la tasa de ocupación.

5.2 Modelo para los años 2008-2020

Como ya se ha comentado anteriormente, se llevarán a cabo distintos modelos a través de los cuales se modelizará la tasa de ocupación con intención de poder analizar no solo el descenso en la tasa de ocupación que supuso el COVID-19, sino que también observar la interrupción del crecimiento que se llevaba observando desde 2013.

En este caso, se dejarán a un lado los datos correspondientes a los años 2005, 2006 y 2007, debido a su lejanía temporal con la crisis del COVID-19. De esta manera el número de tasas de ocupación observadas serán 832 (64 pseudo-individuos x 13 años). Se propondrán dos modelos diferentes para los años 2008-2020, uno análogo al modelo del apartado 5.1, donde la evolución de la tasa de ocupación se modeliza mediante variables ficticias, y otro donde esta evolución se modeliza a través de una parábola, algo que parece razonable a la vista de la figura 11.

5.2.1 Especificación de la tendencia mediante variables ficticias

Modelo para los años 2008-2020 [2]

$$To_{it} = \beta_1 + \beta_2 V_{it} + \beta_3 S_{it} + \beta_4 P_{it} + \beta_5 U_{it} + \beta_6 E1_{it} + \beta_7 E7_{it} + \beta_8 E8_{it} + \beta_9 A08_{it} \\ + \beta_{10} A09_{it} + \beta_{11} A10_{it} + \beta_{12} A11_{it} + \beta_{13} A12_{it} + \beta_{14} A13_{it} + \beta_{15} A14_{it} \\ + \beta_{16} A15_{it} + \beta_{17} A16_{it} + \beta_{18} A17_{it} + \beta_{19} A18_{it} + \beta_{20} A20_{it} + U_{it}$$

$$t = 1, 2, \dots, 13$$

$$i = 1, 2, \dots, 32$$

El modelo [2], al igual que el modelo [1], se ha especificado mediante variables ficticias, creando un modelo análogo al modelo [1], pero para un intervalo temporal menor. Posteriormente se estimó por MCO con la corrección de Arellano a la hora de estimar las desviaciones típicas, obteniendo así las siguientes estimaciones:

Tabla 6: Resultado de las estimaciones del modelo con variables ficticias para los años 2008-2020 para las variables significativas

Variable	Coefficiente	Desv, típica	Estadístico t	Valor p	
const	0,632649	0,0123976	51,03	<0,0001	***
V	0,112402	0,00992185	11,33	<0,0001	***
S	0,114992	0,0156038	7,369	<0,0001	***
P	0,0942930	0,0124013	7,603	<0,0001	***
U	0,217123	0,0118187	18,37	<0,0001	***
E1	-0,120417	0,0262186	-4,593	<0,0001	***
E7	-0,0984795	0,0140624	-7,003	<0,0001	***
E8	-0,421235	0,0161929	-26,01	<0,0001	***
A08	-0,00342372	0,0145626	-0,2351	0,8149	
A09	-0,0385449	0,0121005	-3,185	0,0022	***
A10	-0,0199874	0,0115445	-1,731	0,0883	*
A11	-0,0212301	0,0105464	-2,013	0,0484	**
A12	-0,0570927	0,0101973	-5,599	<0,0001	***
A13	-0,0709866	0,0105383	-6,736	<0,0001	***
A14	-0,0625865	0,00890434	-7,029	<0,0001	***
A15	-0,0460395	0,00972562	-4,734	<0,0001	***
A16	-0,0257934	0,00953999	-2,704	0,0088	***
A17	-0,0145245	0,0110229	-1,318	0,1924	
A18	-0,00215421	0,00872713	-0,2468	0,8058	
A20	-0,0173356	0,0105641	-1,641	0,1058	
<i>R-cuadrado</i>			0,845672		

Fuente: Elaboración propia a partir de los datos de la EPA (INE)

5.2.2 Especificación de la tendencia de forma cuadrática

Para poder evaluar el efecto del COVID se necesita estimar la evolución de la tasa de ocupación en el tiempo mediante algún tipo de función. En este apartado, especificamos un modelo donde la evolución en el tiempo se modeliza mediante una parábola, que en vista de los gráficos 10 y 11 la especificación parece razonable. Es decir, se propone un modelo con tendencia cuadrática y para ello se añadirá la variable ficticia para el año 2020, que recogerá el efecto del COVID y se crearán dos nuevas variables: tiempo y tiempo², entendiendo como tiempo al año correspondiente menos 2019. De esta forma el valor de la tendencia cuadrática tendrá un valor de 0 para el año 2019, convirtiéndose en el año de referencia.

Se propone añadir también otras dos variables: A19 y A20, haciendo referencia a la influencia de cada uno de los años (2019 y 2020) en la tasa de ocupación. Se ha decidido añadir tanto la variable que hace referencia al año del COVID como a la del año anterior ya que como ya se ha comentado anteriormente, los efectos de la pandemia no se han aislado en su totalidad en el año 2020. De hecho, en el año 2019 están incluidas las primeras dos semanas del confinamiento, que aunque se trate de un espacio de tiempo muy breve, sus consecuencias fueron remarcables.

Modelo con tendencia cuadrática para los años 2008-2020 [3]

$$To_{it} = \beta_1 + \beta_2 V_{it} + \beta_3 S_{it} + \beta_4 P_{it} + \beta_5 U_{it} + \beta_6 E1_{it} + \beta_7 E7_{it} + \beta_8 E8_{it} + \beta_9 t_{it} + \beta_{10} t_{it}^2 + \beta_{11} A19 + \beta_{12} A20 + U_{it}$$

$$t = 1, 2, \dots, 13$$

$$i = 1, 2, \dots, 64$$

Estos son los resultados obtenidos tras estimar el modelo [3] por MCO con la corrección de Arellano a la hora de estimar las desviaciones típicas:

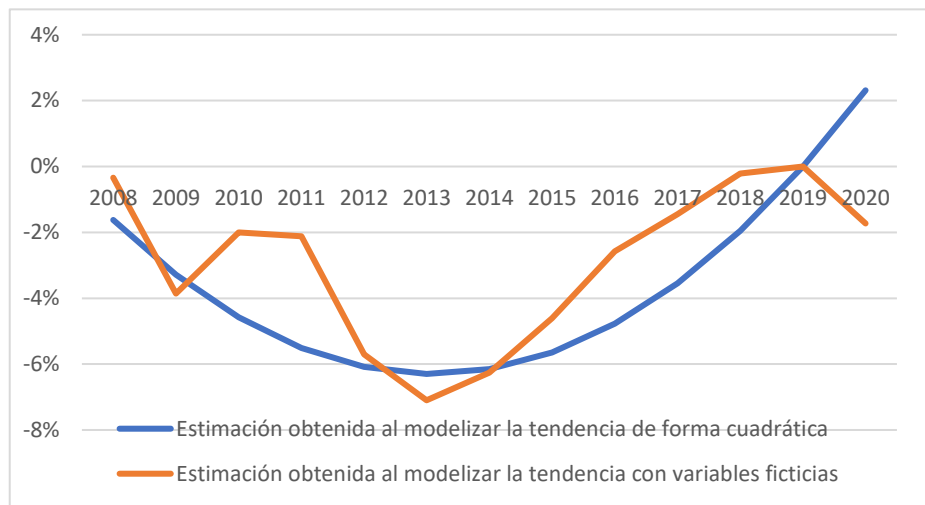
Tabla 7: Resultado de las estimaciones del modelo cuadrático para los años 2008-2020 para las variables significativas

Variable	Coficiente	Desv, típica	Estadístico t	Valor p	
Const	0,656922	0,0145472	45,16	<0,0001	***
V	0,112402	0,00987333	11,38	<0,0001	***
S	0,114992	0,0155275	7,406	<0,0001	***
P	0,0942930	0,0123406	7,641	<0,0001	***
U	0,217123	0,0117609	18,46	<0,0001	***
E1	-0,120417	0,0260904	-4,615	<0,0001	***
E7	-0,0984795	0,0139936	-7,037	<0,0001	***
E8	-0,421235	0,0161138	-26,14	<0,0001	***
tiempo	0,0259263	0,00351083	7,385	<0,0001	***
sq_tiempo	0,00213787	0,000333663	6,407	<0,0001	***
A19	-0,0242733	0,0106289	-2,284	0,0258	**
A20	-0,0696730	0,0182694	-3,814	0,0003	***
R-cuadrado			0,841789		

Fuente: Elaboración propia a partir de los datos de la EPA (INE)

5.2.3 Comparación de la especificación para los años 2008-2020

Figura 12: Estimación de la evolución temporal 2008-2009: variables ficticias VS tendencia cuadrática



Fuente: Elaboración propia a partir de los datos de la EPA (INE)

En los anteriores apartados se han propuesto dos modelos diferentes para el mismo periodo. Como se puede observar en la figura 12, tanto el modelo con variables ficticias para los años 2008-2020 como el modelo con tendencia cuadrática obtienen valores de 0 para el año 2019, el año anterior al COVID-19.

En la especificación del modelo con las variables ficticias para los años 2008-2020 (modelo [2]), el efecto de la pandemia es recogido a través del coeficiente asociado a la variable A20. Como se puede observar en la tabla 4, su valor estimado será de un -1,73%. Como ya se ha comentado anteriormente, en esta especificación no se tiene en cuenta la evolución creciente de la tasa de ocupación en los años posteriores al 2013. La especificación parabólica de la tendencia, junto a las variables A19 y A20 (modelo [3]) trata de solucionar dicho problema.

En esta última especificación se estima que el COVID se reflejó en una caída del 2,4% en el año 2019 y de un 6,9% para el año 2020 (ver tabla 7). Los primeros impactos de la pandemia en el mercado laboral se han contabilizado como parte del año 2019, y es por ello por lo que el efecto es algo elevado.

De esta forma, la variable A20 junto a la tendencia parabólica estimara un mayor efecto del COVID-19 (con un valor del -6,9%) en comparación con la estimación obtenida empleando el modelo de variables ficticias (valor de -1,73%). Esto se debe a que el modelo [3] recoge también el parón de crecimiento y no se limita al descenso de la tasa de ocupación. En cuanto al ajuste, ambos modelos se comportan de forma similar, con un 84,56% y un 84,17%, respectivamente.

5.3 Modelo para los años 2013-2020

5.3.1 Especificación de la tendencia mediante variables ficticias para los años 2013-2020

En el apartado anterior se ha propuesto una parábola para ajustar estimación temporal para incluir la tendencia de crecimiento observada en las figuras 11 y 12. Emplear un modelo parabólico para tratar de predecir un extremo de la muestra puede que no sea lo más acertado, ya que las parábolas tienen unos crecimientos muy acelerados en sus extremos, y a medida que nos alejamos del centro de la parábola la pendiente es aún mayor. Para tratar de solventar este problema se propondrán dos nuevos modelos.

En este caso, teniendo en cuenta que la tasa de ocupación estaba en fase de crecimiento desde el año 2013, suponemos que de no haber sido por el COVID-19, la tasa de ocupación para el año 2020 habría continuado con dicha tendencia. Consideramos que los datos anteriores al 2013 no reflejan dicha tendencia de crecimiento y por ello se dejarán a un lado en este apartado. De esta forma se tratará de estimar la tendencia de forma lineal para los años 2013-2020.

Al igual que en el apartado anterior, se propondrán dos modelos diferentes, uno modelizará la evolución de la tasa de ocupación a través de variables ficticias (modelo [4]) y otro, que en este caso lo hará mediante una función lineal (modelo [5]).

Tanto para el modelo [4] como para el modelo [5] solo se emplearán datos posteriores al 2013. De esta forma el número de tasas de ocupación analizadas serán 512 (64 pseudo-individuos x 8 años). El modelo propuesto para los años 2013-2020 será muy similar al [3] pero para los años posteriores al 2013:

Modelo para los años 2013-2020 [4]

$$\begin{aligned}
 To_{it} = & \beta_1 + \beta_2 V_{it} + \beta_3 S_{it} + \beta_4 P_{it} + \beta_5 U_{it} + \beta_6 E1_{it} + \beta_7 E7_{it} + \beta_8 E8_{it} + \beta_9 A13_{it} \\
 & + \beta_{10} A14_{it} + \beta_{11} A15_{it} + \beta_{14} A16_{it} + \beta_{15} A17_{it} + \beta_{16} A18_{it} + \beta_{17} A20_{it} + U_{it}
 \end{aligned}$$

$$\begin{aligned}
 t &= 1, 2, \dots, 8 \\
 i &= 1, 2, \dots, 64
 \end{aligned}$$

El modelo [4], al igual que los anteriores modelos, se ha estimado por MCO con la corrección de Arellano a la hora de estimar las desviaciones típicas. Las estimaciones obtenidas se reflejan en la siguiente tabla.

Tabla 8: Resultado de las estimaciones del modelo con variables ficticias para los años 20013-2020 para las variables significativas

Variable	Coefficiente	Desv. típica	Estadístico t	Valor p	
Const	0,631563	0,0131024	48,20	<0,0001	***
V	0,0960074	0,0103980	9,233	<0,0001	***
S	0,124050	0,0153254	8,094	<0,0001	***
P	0,0948462	0,0144795	6,550	<0,0001	***
U	0,228973	0,0115511	19,82	<0,0001	***
E1	-0,140194	0,0293011	-4,785	<0,0001	***
E7	-0,0787015	0,0109954	-7,158	<0,0001	***
E8	-0,389891	0,0163246	-23,88	<0,0001	***
A13	-0,0709866	0,0105628	-6,720	<0,0001	***
A14	-0,0625865	0,00892506	-7,012	<0,0001	***
A15	-0,0460395	0,00974826	-4,723	<0,0001	***
A16	-0,0257934	0,00956219	-2,697	0,0090	***
A17	-0,0145245	0,0110486	-1,315	0,1934	
A18	-0,00215421	0,00874744	-0,2463	0,8063	
A20	-0,0173356	0,0105887	-1,637	0,1066	
R-cuadrado			0,845869		

Fuente: Elaboración propia a partir de los datos de la EPA (INE)

5.3.2 Especificación de la tendencia de forma lineal

Tal y como se ha comentado en la sección 5.2, en los modelos [1], [2] y [4], no se tiene en cuenta la inercia de crecimiento que tenía la tasa de ocupación en ese año. Esto es lo que se trata de corregir con las especificaciones de los modelos [3] y [5].

A través del modelo [5], al igual que con el modelo [3], se intentará estimar no solo el descenso que supuso el 2020, sino también el parón de crecimiento. Por ello, con intención de acercarnos en mayor medida a la realidad, se estimará una tendencia lineal para los años posteriores al 2013, obteniendo así el modelo [5]. Al igual que en el modelo [3], se incluirán las variables A19 y A20.

Modelo para los años 20013-2020 [5]

$$T0_{it} = \beta_1 + \beta_2 V_{it} + \beta_3 S_{it} + \beta_4 P_{it} + \beta_5 U_{it} + \beta_6 E1_{it} + \beta_7 E7_{it} + \beta_8 E8_{it} + \beta_9 t_{it} + \beta_{10} A19 + \beta_{11} A20 + U_{it}$$

$$t = 1, 2, \dots, 8$$

$$i = 1, 2, \dots, 64$$

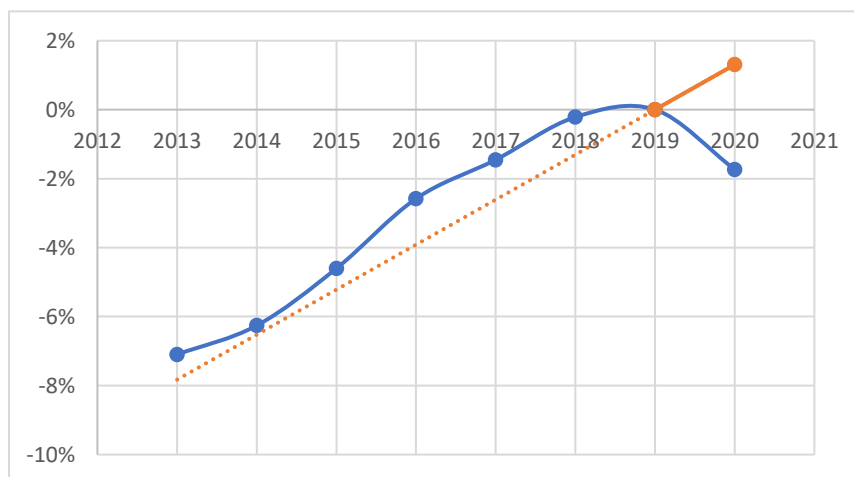
Tabla 9: Resultado de las estimaciones del modelo lineal para los años 20013-2020 para las variables representativas

Variable	Coefficiente	Desv. Típica	Estadístico t	Valor p	
Cons	0,645408	0,0147934	43,63	<0,0001	***
V	0,0960074	0,0103565	9,270	<0,0001	***
S	0,124050	0,0152641	8,127	<0,0001	***
P	0,0948462	0,0144216	6,577	<0,0001	***
U	0,228973	0,0115049	19,90	<0,0001	***
E1	-0,140194	0,0291839	-4,804	<0,0001	***
E7	-0,0787015	0,0109515	-7,186	<0,0001	***
E8	-0,389891	0,0162593	-23,98	<0,0001	***
tiempo	0,0145313	0,00186297	7,800	<0,0001	***
A19	-0,0138453	0,00972573	-1,424	0,1595	
A20	-0,0457122	0,0152507	-2,997	0,0039	***
R-cuadrado			0,845702		

Fuente: Elaboración propia a partir de los datos de la EPA (INE)

5.3.3 Comparación de la especificación para los años 2013-2020

Figura 13: Estimación de la evolución temporal: 2013-2020: variables ficticias VS tendencia lineal



Fuente: Elaboración propia a partir de los datos de la EPA (INE)

Al enfrentar la estimación obtenida con las variables ficticias y la obtenida al estimar la tendencia lineal, al igual que en la sección 5.2.3, se obtendrán valores de 0 para el año 2019, por lo que los valores recogidos para el año 2020 reflejarán la influencia del COVID-19. Cabe destacar que los resultados obtenidos para el año 2020 a través del modelo estimado mediante variables ficticias,

no reflejará la tendencia de crecimiento que se veía desde 2013, mientras que la estimada de forma lineal sí.

La especificación del modelo con las variables ficticias para los años 2013-2020 (modelo [4]), recoge el efecto del COVID-19 en la tasa de ocupación a través del coeficiente asociado a la variable A20, en este caso un valor de -0,0173356. Esto sugiere que la pandemia ha generado una tasa de ocupación un 1,73% menor que la tasa prepandemia.

Por otro lado, el modelo [5] sugiere que el efecto del COVID-19 supuso un descenso de la tasa de ocupación del 4,57% en el 2020 y del 1,38% en el 2019. Cabe mencionar que como ya se ha comentado anteriormente el primer trimestre de 2020 estará incluido en 2019, y de no ser así el efecto del 2019 será más leve.

Es decir, al igual que en el apartado anterior, la tendencia lineal estimará un mayor efecto de la pandemia en la tasa de ocupación que la estimación del modelo de variables ficticias, debido a que el primer modelo recogerá el parón de crecimiento de la tasa de ocupación, mientras que el modelo estimado a través de variables ficticias se limitará únicamente a reflejar el descenso de la tasa.

5.4 Comparación de modelos estimados

Tabla 10: Comparación de los diferentes modelos

Modelos	Modelo variables ficticias 2005-2020 [1]	Modelo variables ficticias 2008-2020 [2]	Modelo tendencia parabólica 2008-2020 [3]	Modelo variables ficticias 2013-2020 [4]	Modelo tendencia lineal 2013-2020 [5]
Cons	0,628252	0,632649	0,656922	0,631563	0,645408
V	0,130450	0,112402	0,112402	0,0960074	0,0960074
S	0,108910	0,114992	0,114992	0,124050	0,124050
P	0,0958507	0,0942930	0,0942930	0,0948462	0,0948462
U	0,208747	0,217123	0,217123	0,228973	0,228973
E1	-0,105655	-0,120417	-0,120417	-0,140194	-0,140194
E7	-0,115190	-0,0984795	-0,0984795	-0,0787015	-0,0787015
E8	-0,430504	-0,421235	-0,421235	-0,389891	-0,389891
A19			-0,0242733		-0,0138453
A20	-0,0173356	-0,0173356	-0,0696730	-0,0173356	-0,0457122
R ²	0,834613	0,845672	0,841789	0,845869	0,845702

Fuente: Elaboración propia a partir de los datos de la EPA (INE)

En la tabla 10 aparecen resumidos los resultados de los modelos estimados a lo largo del apartado 5. Al analizar las estimaciones obtenidas podemos observar varios fenómenos.



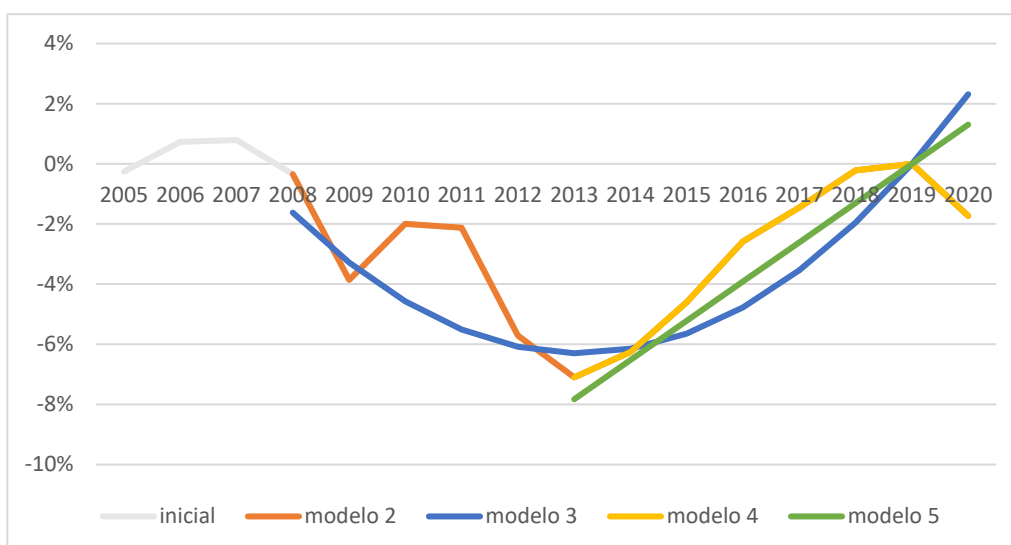
En el modelo inicial, modelo que incluye datos para el periodo 2005-2020, estima que los varones obtienen una tasa de ocupación un 13,04 puntos porcentuales mayor que las de mujeres. El mismo método de estimación empleado para el periodo 2013-2020 obtiene que los varones tienen una ocupación un 9,6% mayor. Esta disminución puede ser consecuencia de la disminución de la brecha de género a lo largo de los años.

Los resultados de la estimación del efecto de la formación en la tasa de ocupación son bastante homogéneos en las distintas especificaciones. Todas las estimaciones concluyen que quienes dispongan únicamente de un nivel de estudios básico serán quienes presenten de tasas de ocupación inferiores, y quienes dispongan de estudios superiores tendrán las tasas más elevadas, esta influencia varía ligeramente según el periodo analizado. Por ejemplo se estima que para el periodo 2005-2020 las tasas de ocupación para quienes dispongan de orientación general obtendrán una tasa 10,89% mayor que quienes únicamente dispongan de orientación básica, pero para el periodo 2013-2020 la diferencia entre ambas tasas será de un 12,4%.

En el caso del grupo de edad, se puede concluir que el grupo de edad 30-54 años será el grupo con mayores tasas de ocupación, independientemente del periodo analizado. Cabe destacar que las tasas de ocupación para el grupo de edad 25-29 años con los años han ido aumentando su diferencia respecto a las del grupo de edad 30-54, obteniendo unas tasas 10,56 puntos porcentuales inferiores para el periodo 2005-2020, mientras que para el periodo 2013-2020 dicha disminución sería de 14 puntos porcentuales. Por otro lado en el grupo de edad entre 60 y 64 años ocurre el caso contrario: a lo largo de los años la diferencia entre las tasas de ocupación para dicho grupo y el grupo de referencia ha ido disminuyendo. Mientras que las tasas de ocupación estimadas para el grupo de edades entre 60 y 64 años para el periodo 2005-2020 es de un 43% inferior al del grupo de edad 30-50, en el periodo 2013-2020 es de un 38,9% inferior. Con estos datos expuestos podemos suponer que cada vez los individuos se incorporan y se retiran a una mayor edad al mercado laboral.

Los modelos estimados a través de las variables ficticias para los diferentes años obtienen resultados bastante similares en cuanto a la evolución temporal. Los valores estimados referidos al efecto del COVID-19 en la tasa de ocupación se pueden dividir en tres diferentes en base al método de especificación: (a) Efecto estimado especificando la tendencia mediante variables ficticias, (b) Efecto estimado de forma cuadrática para los años 2008-2020 y (c) Efecto estimado a través de la especificación de tendencia de forma lineal para los los 2013-2020.

Figura 14: Comparación de los diferentes modelos estimados



Fuente: Elaboración propia a partir de los datos de la EPA (INE)

Como se puede observar en el gráfico 14, basándonos en el método de especificación de tendencia mediante variables ficticias los efectos del COVID habrían sido menores que los efectos estimados en los otros dos métodos. Esto se debe a que como ya se ha comentado en el apartado 5, especificando la tendencia de forma cuadrática o de forma lineal se logra integrar también la disminución de crecimiento, y no únicamente el descenso generado.

De esta forma, la estimación del efecto del COVID-19 en la tasa de ocupación de la CAV variará según el método de estimación empleado. Por un lado, aplicando el método de especificación mediante variables ficticias se considerará que el efecto de la pandemia supuso un descenso en el 2020 de alrededor del 1,73%. Podríamos considerar este 1,73% como un extremo inferior del efecto del COVID en la tasa de ocupación. Por otro lado, las estimaciones obtenidas con la modelización temporal cuadrática y lineal concluyen que continuando con la inercia de crecimiento anterior al 2020, el COVID habría generado una disminución del 6,9% y 4,57% respectivamente. Esta diferencia se debe a que las tendencias cuadráticas tienden a crecer a un ritmo mucho más elevado que las lineales en los extremos, especialmente la parábola. Ambas estimaciones, la cuadrática y la lineal, obtienen valores más elevados que las obtenidas con las variables ficticias debido a que estas también analizan la inercia de crecimiento.

Cabe destacar que al incluir el primer trimestre de 2020 en el año 2019, parece que no se ha conseguido aislar totalmente el efecto del COVID, ya que la segunda quincena del mes de marzo, cuando comenzó el confinamiento, pertenecerá al año anterior. Aunque el COVID únicamente estuviera presente en una octava parte de aquel trimestre que hemos incluido en el año 2019, fueron las dos primeras semanas de la pandemia, semanas que sufrieron un gran descenso en la tasa de ocupación. De hecho, aunque de aquel trimestre únicamente el COVID

estuviera presente únicamente en $1/8$ parte, los efectos correspondientes a dichas semanas fueron muy elevados, debido a que los primeros días del confinamiento se produjeron muchos cambios en el mundo laboral. Podría ser posible que el efecto del COVID fuera aún mayor y por ello en los modelos [3] y [5] se ha incluido la variable A19.

El R^2 , también conocido como coeficiente de determinación, refleja en qué porcentaje de la variabilidad de la tasa de ocupación es explicada con la regresión. El coeficiente estimado del R-cuadrado de todos los modelos obtienen unos valores similares en torno al 85%, valor que se puede considerar como elevado. Además, se puede considerar que las tendencias cuadráticas y lineales se ajustan bastante bien, ya que el R-cuadrado para los diferentes modelos no es demasiado diferente entre sí, y al pasar de variables ficticias a tendencias lineales o cuadráticas no se pierde demasiado en términos de ajuste.

5.5 Interpretación de los coeficientes

A lo largo de este TFG se han propuesto cinco modelos diferentes, y se deberá seleccionar el más adecuado. Cabe destacar que las especificaciones dan resultados similares en cuanto al efecto del sexo, edad y formación en la tasa de ocupación.

Por otro lado, el efecto del COVID queda mejor estimado en los modelos [3] y [5], ya que reflejan también la inercia de crecimiento. A la hora de seleccionar un único modelo, elegiremos el [5], consideramos que la tendencia cuadrática del modelo [3] crecería excesivamente en el año 2020 y la especificación de la tendencia de forma lineal se ajustará mejor a la realidad.

Modelo estimado para los años 2013-2020 [5]

$$\widehat{T}O_{it} = 0,645408 + 0,0960074 V_{it} + 0,124050 S_{it} + 0,0948462 P_{it} + 0,228973 U_{it} \\ - 0,140194 E1_{it} - 0,0787015 E7_{it} - 0,389891 E8_{it} + 0,0145313 t_{it} \\ - 0,0138453 A19 - 0,0457122 A20$$

$$t = 1, 2, \dots, 8$$

$$i = 1, 2, \dots, 64$$

El individuo de referencia se trata de una mujer con estudios obligatorios, con una edad comprendida entre 30 y 54 años en el año 2019. A continuación, se interpretarán cada uno de los coeficientes obtenidos.

- $\widehat{\beta}_1 = 0,645408$. Tasa de ocupación media estimada de un individuo de referencia. En este caso, la tasa de ocupación media en el año 2019, para una mujer de edad entre 30 y 54 años, con un nivel de estudios de educación obligatoria se estima que es del 64,54%.
- $\widehat{\beta}_2 = 0,0960074$. Diferencia entre la tasa de ocupación de un varón con respecto a una mujer manteniendo constantes el resto de regresores, es decir, con la misma formación,

perteneciendo al mismo grupo de edad y en el mismo año. Se estima que la tasa de ocupación media para un varón es un 9,6% superior que la de una mujer con mismas características de formación, edad y año.

- $\widehat{\beta}_3 = 0,124050$. Diferencia entre la tasa de ocupación de un individuo con educación de orientación general frente a uno con formación obligatoria, ceteris paribus. Se estima que la tasa de ocupación media sea un 12,4% superior que la del individuo de mismo sexo, edad y año.
- $\widehat{\beta}_4 = 0,0948462$. Diferencia entre las tasas de ocupación de un individuo con educación de orientación profesional frente a uno con formación obligatoria. Se estima que en 2019 la tasa de ocupación media para un individuo con estudios de orientación profesional sea un 9,48% superior que la del individuo con formación obligatoria y mismas características de sexo, edad y año.
- $\widehat{\beta}_5 = 0,228973$. Diferencia entre las tasas de ocupación de un individuo con educación superior frente a uno con educación obligatoria, ceteris paribus. Se estima que la tasa de ocupación para un individuo con estudios de educación superior sea un 22,89% superior que la del individuo con formación obligatoria y mismas características de sexo, edad y año.
- $\widehat{\beta}_6 = -0,140194$. Diferencia entre las tasas de ocupación de un individuo con una edad entre 25 y 29 frente a uno de edad comprendida entre 30 y 54 años, ceteris paribus. Se estima que la tasa de ocupación media para un individuo de edad entre 25 y 29 años sea un 14 % inferior que la de un individuo con una edad entre 30 y 54 años y con mismas características de sexo, formación y año.
- $\widehat{\beta}_7 = -0,0787015$. Diferencia entre las tasas de ocupación de un individuo con una edad entre 55 y 59 frente a uno de edad entre 30 y 54 años, ceteris paribus. Se estima que la tasa de ocupación media para un individuo de edad entre 55 y 59 años sea un 7,8 % inferior que la de un individuo con una edad entre 30 y 54 años y mismas características de sexo, formación y año.
- $\widehat{\beta}_8 = -0,389891$. Diferencia entre las tasas de ocupación de un individuo con una edad entre 60 y 64 frente a uno de edad entre 30 y 54 años, ceteris paribus. Se estima que la tasa de ocupación media para un individuo de edad entre 60 y 64 años sea un 38,94 % inferior que la del individuo con una edad entre 30 y 54 años, con mismas características de sexo, formación y año.
- $\widehat{\beta}_9 = 0,0145313$. Estima el efecto del tiempo en la tasa de ocupación. En el periodo muestral analizado (2013-2020), por cada año que pasa se estima que la tasa de ocupación aumenta en un 1,45%, manteniendo el resto de los regresores constantes.
- $\widehat{\beta}_{10} = -0,0138453$. Estima la diferencia en la tasa de ocupación del año 2019 con la que se habría dado en ese año si se hubiese seguido con el crecimiento lineal propuesto en el modelo. En este caso se estima que la tasa de ocupación en 2019 tuvo un descenso del 1,38%.
- $\widehat{\beta}_{11} = -0,0457122$. Estima la diferencia entre las tasas de ocupación de 2020 respecto a la de 2019, después de haber descontado la caída extraordinaria del 2019. Se estima que



en este caso el descenso de la tasa de ocupación en 2020 provocado por la pandemia fue de un 4,5%.

6. CONCLUSIONES

El objetivo de este análisis es poder estimar en qué medida ha influido la pandemia del COVID-19 en la tasa de ocupación de la CAV, al igual que otras variables como podrían ser el sexo, nivel de formación o edad. En el modelo [5], se estima que el efecto del coronavirus ha generado una caída de la tasa de ocupación del 1,38% en el periodo del 2º trimestre del 2019 a 1º trimestre de 2020 y de una caída del 4,5% del 2º trimestre del 2020 al primer trimestre de 2021.

Aunque la influencia que ha tenido el COVID en la tasa de ocupación haya sido importante, existen factores que tienen una mayor influencia en la tasa de ocupación. Por ejemplo, mientras que la pandemia ha generado una disminución del 4,5% en la tasa de ocupación, hay factores como la edad, que tienen una mayor repercusión, donde el grupo de edad más joven obtendrá tasas de ocupación medias un 14% inferiores, una influencia casi 10 puntos porcentuales mayor que la del COVID-19. De hecho, entre las variables explicativas analizadas, es el COVID-19 la variable que menor repercusión ha tenido en el mercado laboral.

Respecto a la influencia que tiene el sexo en la tasa de ocupación, a pesar de que con los años el papel de la mujer haya ido evolucionando y se haya ido incorporando al mercado laboral, a día de hoy todavía soportan tasas de ocupación inferiores a la de los hombres. Se estima que la tasas de ocupación para los varones sean 9,6% superiores siempre y cuando el resto de condicionantes sean constantes (sexo, edad y formación). De hecho, como ya se ha comentado en el apartado del análisis descriptivo del sexo, la brecha de género en el mercado laboral ha ido disminuyendo, pero desde la crisis financiera del 2008 se ha mantenido relativamente constante.

Por otro lado, los grupos de edad de los extremos (tanto los que se están incorporando al mercado laboral como los que están comenzando a retirarse) son los que menores tasas de ocupación representan, mientras que el grupo de edad entre 30 y 54 años será el que mayores tasas obtenga. Por otro lado, los grupos de edad quinquenales entre 30 y 54 años no son relevantes y todos obtienen unas tasas similares entre sí, y por ello se han agrupado en este estudio.

Por último, quienes dispongan de estudios de formación obligatoria serán quienes menores tasas de ocupación preseten y las tasas de ocupación serán mayores entre los que tengan estudios de educación superior. Este factor no es de extrañar debido a que cada vez las generaciones están más preparadas y como se ha podido observar en la figura 7, cada vez existe una menor proporción de individuos con únicamente educación obligatoria mientras que quienes disponen de educación superior han ido en aumento desde el 2005.

En definitiva, se puede concluir que la mayor tasa de ocupación media esperada pertenecerá a un varón de edad comprendida entre 35 y 54 años, con estudios de formación superior, mientras que los jóvenes, las mujeres y quienes dispongan únicamente de estudios de formación obligatoria tendrán menores tasas de empleo. Cabe destacar que dichos grupos poblacionales



tienden a tener mayores tasas de temporalidad, por lo que sufrirán en mayor medida los efectos de las crisis, y mas si en dichas crisis, como en la originada por la pandemia, sufren altos niveles de destrucción de empleo en sectores como la hostelería y servicios, dónde la precariedad y temporalidad es aún mayor.

Aunque el proposito de este trabajo no es analizar el por qué de las tasas de incidencia de la pandemia en la tasa de ocupación que se obtenido, parece razonable pensar que si la influencia del covid en el mercado laboral ha sido tan baja, se debe principalmente a la cobertura social de los ERTE, que impidió la destrucción masiva de empleo, y la posibilidad que permite la tecnología actual de teletrabajar. Sin estos dos elementos, los datos obtenidos habrían sido diferentes.

7. REFERENCIAS

- Alonso, A., Fernández-Macho, F. J. & Gallastegui, I. (2005). *Econometría*, Madrid: Pearson Educación, S.A.
- Coll, F. & López, J. F. (2020). Los ERTE en cifras, *Fundación Civismo 08/2020*, 1-17.
- Deaton, A. (1985). Panel data from time series of cross-sections, *Journal of econometrics*, 30, 109-126.
- Dirección de Economía y Planificación, (2020). *Informe anual de la economía vasca 2020*, Departamento de Economía y Hacienda, Donostia-San Sebastián.
- Gómez, M., Hospido, L., & Sanz, C. (2021). El impacto diferencial por sexos de la crisis del COVID-19 en el mercado de trabajo, *Boletín económico/Banco de España*, 3, 39-42.
- Gómez, R. & García Perea, P. (1994). Elaboración de series históricas de empleo a partir de la Encuesta de Población Activa (1964-1992), Documentos de trabajo / Banco de España, 9409,
- Greenwood A. M. (1999). Definiciones internacionales y futuro de las estadísticas del Subempleo , Oficina de Estadística de la OIT (Oficina Internacional del Trabajo), Ginebra.
- Gutiérrez, E. & Moral-Benito, E. (2020). Medidas de contención, evolución del empleo y la propagación del COVID -19 en los municipios Españoles, *Documentos ocasionales / Banco de España*, 2022.
- Instituto Nacional de Estadística, (2021). *Encuesta de población activa: metodología 2021*, Madrid.
- López, J. F. (2020). Jóvenes españoles, una generación en crisis, *Fundación Civismo*, 1-30.
- Mackay-Castro, C. R., León, B. V. & Zambrano-Noboa, H. A. (2020). Efectos del Covid-19 en el mercado laboral, *Dominio de las ciencias*, 6(3), 1368-1381.
- Mahia, R. (2021). Los efectos del COVID-19 sobre la inmigración en España: economía, trabajo y condiciones de vida, *Anuario CIDOB de la inmigración 2020*, 68-81.
- Medina, L. G. (2004). Demografía longitudinal de la ocupación, *Información Comercial Española*, 815, 105-112.
- Oficina Internacional del Trabajo, (2-11 de Octubre de 2013), Estadísticas del trabajo, el empleo y la subutilización de la fuerza de trabajo, *19,ª Conferencia Internacional de Estadísticos del Trabajo*, Ginebra.



- Oguiza, A. Gallastegui, I. & Núñez-Antón, V. (2012), Analysis of pseudo-panel data with dependent samples, *Journal of Applied Statistics*, 39, 1921-1937.
- Real Decreto 463/2020, de 14 de marzo, por el que se declara el estado de alarma para la gestión de la situación de crisis sanitaria ocasionada por el COVID-19, 67 (Ministerio de la Presidencia, Relaciones con las Cortes y Memoria Democrática 2020).
- Real Decreto-ley 8/2020, de 17 de marzo, de medidas urgentes extraordinarias para hacer frente al impacto económico y social del COVID-19., 73 (Jefatura de estado 2020).
- Rocha, F. (2012). La crisis económica y sus efectos sobre el empleo en España, *Gaceta sindical*, 19, 69-89.
- Roncaglia, A. (2006). Tasa de desempleo y tasas de empleo: ¿categorías estadísticas o construcciones teóricas? *Investigación Económica*, 65, 46-61.
- Ruesga, S. & Viñas, A. (2020). Desempleo y ERTes: un dilema para España ante la pandemia de COVID-19, *Economía UNAM*, 18 (52). 87-106.
- Torres, A. (2020). Estudio y análisis de resultados de la Encuesta de Población Activa (EPA), TFG de la Facultad de ciencias sociales y jurídicas de Jaén,
- Villagarcía, T. (1995). *¿Existe un sesgo de inactividad en la encuesta de población activa?* Documento de trabajo 94-04, Serie de estadística y Econometría, Universidad Carlos III de Madrid.
- Yeregui, A. & Vallejo, L. (2021). El rastreo de la COVID-19 en el mundo laboral, La experiencia del País Vasco, *Revista Española de Salud Pública*, 95, 1-16.

8. APÉNDICE

8.1 Modelo para el periodo 2008-2020 incluyendo todas las variables (incluso las no significativas)

Modelo con variables ficticias para los años 2008-2020 [7] para todas las variables

$$\begin{aligned}
 To_{it} = & \beta_1 + \beta_2 V_{it} + \beta_3 S_{it} + \beta_4 P_{it} + \beta_5 U_{it} + \beta_6 E1_{it} + \beta_7 E2_{it} + \beta_8 E3_{it} + \beta_9 E5_{it} \\
 & + \beta_{10} E6_{it} + \beta_{11} E7_{it} + \beta_{12} E8_{it} + \beta_{13} A08_{it} + \beta_{14} A09_{it} + \beta_{15} A10_{it} \\
 & + \beta_{16} A11_{it} + \beta_{17} A12_{it} + \beta_{18} A13_{it} + \beta_{19} A14_{it} + \beta_{20} A15_{it} + \beta_{21} A16_{it} \\
 & + \beta_{22} A17_{it} + \beta_{23} A18_{it} + \beta_{24} A20_{it} + U_{it}
 \end{aligned}$$

$$t = 1, 2, \dots, 13$$

$$i = 1, 2, \dots, 64$$

Tabla 11: Resultado de las estimaciones del modelo para los años 2008-2020 para todas las variables

Variable	Coficiente	Desv. típica	Estadístico t	Valor p	
const	0,640581	0,0130626	49,04	6,72e-052	***
V	0,112402	0,00978024	11,49	4,16e-017	***
S	0,114992	0,0154570	7,439	3,41e-010	***
P	0,0942930	0,0118503	7,957	4,24e-011	***
U	0,217123	0,0112540	19,29	3,92e-028	***
E1	-0,128349	0,0268420	-4,782	1,08e-05	***
E2	-0,0208205	0,0135854	-1,533	0,1304	
E3	0,000133983	0,0106490	0,01258	0,9900	
E5	-0,00232482	0,00893302	-0,2602	0,7955	
E6	-0,0166497	0,0143484	-1,160	0,2503	
E7	-0,106412	0,0151130	-7,041	1,69e-09	***
E8	-0,429167	0,0171226	-25,06	1,79e-034	***
A08	-0,00342372	0,0145986	-0,2345	0,8153	
A09	-0,0385449	0,0121304	-3,178	0,0023	***
A10	-0,0199874	0,0115731	-1,727	0,0891	*
A11	-0,0212301	0,0105725	-2,008	0,0489	**
A12	-0,0570927	0,0102225	-5,585	5,32e-07	***
A13	-0,0709866	0,0105643	-6,719	6,15e-09	***
A14	-0,0625865	0,00892635	-7,011	1,91e-09	***

A15	-0,0460395	0,00974967	-4,722	1,35e-05	***
A16	-0,0257934	0,00956357	-2,697	0,0090	***
A17	-0,0145245	0,0110502	-1,314	0,1935	
A18	-0,00215421	0,00874870	-0,2462	0,8063	
A20	-0,0173356	0,0105902	-1,637	0,1066	
R-cuadrado			0,847165		

Fuente: Elaboración propia a partir de los datos de la EPA (INE)

Modelo con tendencia cuadrática para los años 2008-2020 [8] para todas las variables

$$\begin{aligned}
 T_{oit} = & \beta_1 + \beta_2 V_{it} + \beta_3 S_{it} + \beta_4 P_{it} + \beta_5 U_{it} + \beta_6 E1_{it} + \beta_7 E2_{it} + \beta_8 E3_{it} + \beta_9 E5_{it} \\
 & + \beta_{10} E6_{it} + \beta_{11} E7_{it} + \beta_{12} E8_{it} + \beta_{13} t + \beta_{14} t^2 + \beta_{15} A20 + U_{it} \\
 & t = 1, 2, \dots, 13 \\
 & i = 1, 2, \dots, 64
 \end{aligned}$$

Tabla 12: Resultado de las estimaciones del modelo cuadrático para los años 2008-2020 para todas las variables

Variable	Coeficiente	Desv. típica	Estadístico t	Valor p	
Const	0,651584	0,0126997	51,31	4,17e-053	***
V	0,112402	0,00972622	11,56	3,28e-017	***
S	0,114992	0,0153717	7,481	2,89e-010	***
P	0,0942930	0,0117848	8,001	3,55e-011	***
U	0,217123	0,0111919	19,40	2,91e-028	***
E1	-0,128349	0,0266938	-4,808	9,84e-06	***
E2	-0,0208205	0,0135104	-1,541	0,1283	
E3	0,000133983	0,0105902	0,01265	0,9899	
E5	-0,00232482	0,00888368	-0,2617	0,7944	
E6	-0,0166497	0,0142692	-1,167	0,2477	
E7	-0,106412	0,0150296	-7,080	1,45e-09	***
E8	-0,429167	0,0170280	-25,20	1,30e-034	***
tiempo	0,0213250	0,00253478	8,413	6,78e-012	***
sq_tiempo	0,00180445	0,000263940	6,837	3,85e-09	***
A20	-0,0514680	0,0131904	-3,902	0,0002	***
R-cuadrado			0,842671		

Fuente: Elaboración propia a partir de los datos de la EPA (INE)

8.2 Modelo para el periodo 2013-2020 incluyendo todas las variables (incluso las no significativas)

Modelo para los años 2013-2020 [9]

$$\begin{aligned}
 T o_{it} = & \beta_1 + \beta_2 V_{it} + \beta_3 S_{it} + \beta_4 P_{it} + \beta_5 U_{it} + \beta_6 E1_{it} + \beta_7 E2_{it} + \beta_8 E3_{it} + \beta_9 E5_{it} \\
 & + \beta_{10} E6_{it} + \beta_{11} E7_{it} + \beta_{12} E8_{it} + \beta_{13} A13_{it} + \beta_{14} A14_{it} + \beta_{15} A15_{it} \\
 & + \beta_{16} A16_{it} + \beta_{17} A17_{it} + \beta_{18} A18_{it} + \beta_{19} A20_{it} + U_{it}
 \end{aligned}$$

$$t = 1, 2, \dots, 8$$

$$i = 1, 2, \dots, 64$$

Tabla 13: Resultado de las estimaciones del modelo con variables ficticias para los años 20013-2020 para todas las variables

Variable	Coficiente	Desv. típica	Estadístico t	Valor p	
Cons	0,637753	0,0140569	45,37	7,90e-050	***
V	0,0960074	0,0102840	9,336	1,71e-013	***
S	0,124050	0,0153668	8,073	2,66e-011	***
P	0,0948462	0,0139104	6,818	4,14e-09	***
U	0,228973	0,0108359	21,13	2,70e-030	***
E1	-0,146384	0,0301533	-4,855	8,30e-06	***
E2	-0,0212765	0,0163125	-1,304	0,1969	
E3	0,00124288	0,0121076	0,1027	0,9186	
E5	0,000502798	0,0103204	0,04872	0,9613	
E6	-0,0114210	0,0134568	-0,8487	0,3993	
E7	-0,0848918	0,0128679	-6,597	1,00e-08	***
E8	-0,396082	0,0176736	-22,41	1,02e-031	***
A13	-0,0709866	0,0106056	-6,693	6,82e-09	***
A14	-0,0625865	0,00896120	-6,984	2,13e-09	***
A15	-0,0460395	0,00978773	-4,704	1,44e-05	***
A16	-0,0257934	0,0096009	-2,687	0,0092	***
A17	-0,0145245	0,0110933	-1,309	0,195	
A18	-0,00215421	0,00878286	-0,2453	0,8070	
A20	-0,0173356	0,0106316	-1,631	0,1080	
R-cuadrado			0,847466		

Fuente: Elaboración propia a partir de los datos de la EPA (INE)

Modelo para los años 20013-2020 [10]

$$T o_{it} = \beta_1 + \beta_2 V_{it} + \beta_3 S_{it} + \beta_4 P_{it} + \beta_5 U_{it} + \beta_6 E1_{it} + \beta_7 E2_{it} + \beta_8 E3_{it} + \beta_9 E5_{it} + \beta_{10} E6_{it} + \beta_{11} E7_{it} + \beta_{12} E8_{it} + \beta_{13} t + \beta_{14} A20 + U_{it}$$

$$t = 1, 2, \dots, 8$$

$$i = 1, 2, \dots, 64$$

Tabla 14: Resultado de las estimaciones del modelo lineal para los años 2013-2020 para todas las variables

Variable	Coficiente	Desv. Típica	Estadístico t	Valor p	
Cons	0,645170	0,0138489	46,59	1,56e-050	***
V	0,0960074	0,0102323	9,383	1,42e-013	***
S	0,124050	0,0152894	8,113	2,26e-011	***
P	0,0948462	0,0138403	6,853	3,60e-09	***
U	0,228973	0,0107813	21,24	2,04e-030	***
E1	-0,146384	0,0300016	-4,879	7,59e-06	***
E2	-0,0212765	0,0162305	-1,311	0,1947	
E3	0,00124288	0,0120467	0,1032	0,9182	
E5	0,000502798	0,0102684	0,04897	0,9611	
E6	-0,0114210	0,0133891	-0,8530	0,3969	
E7	-0,0848918	0,0128032	-6,631	8,76e-09	***
E8	-0,396082	0,0175847	-22,52	7,69e-032	***
tiempo	0,0130478	0,00144260	9,045	5,42e-013	***
A20	-0,0378006	0,0120789	-3,129	0,0027	***
R-cuadrado			0,846882		

Fuente: Elaboración propia a partir de los datos de la EPA (INE)