



Universidad
Zaragoza

Trabajo Fin de Grado

Diferencias salariales entre hombres y mujeres en
España durante el periodo de recesión entre 2008 y
2014

Wage differences between men and women in Spain
during recession from 2008 to 2014

Autor/es

Cristian Farapon

Director/es

Jesús Serafín Clemente López

Monia Ben-Kaabia

Facultad de Economía y Empresa

2017/2018

RESUMEN

El trabajo se enfoca en el mercado laboral español para estudiar la evolución de la diferencia salarial entre hombres y mujeres en España durante la Gran Recesión primero a nivel nacional y seguidamente entre comunidades autónomas de Aragón, País Vasco y Andalucía utilizando variables descriptivas del mercado laboral. Como señala García (2017) las diferencias en los indicadores laborales cambian con el ciclo económico, y por tanto un periodo de recesión puede influir en la brecha por género. Se observa que las diferencias en paro se reducen durante la recesión y también que el tamaño de las diferencias en salarios difiere si hablamos de contratos indefinidos y contratos temporales. Por último, construimos un modelo econométrico involucrando 17 comunidades autónomas, exceptuando Ceuta y Melilla, utilizando la técnica de datos de panel. En consecuencia, este trabajo pretende analizar la evolución de la brecha salarial por género durante la actual crisis desde una perspectiva regional, poniendo el énfasis en la capacidad de explicar la evolución de dicha brecha de la diferencia en tasas de paro y en la estructura productiva de las regiones.

ABSTRACT

The work focuses on the Spanish labor market to study the evolution of the salary difference between men and women in Spain during the Great Recession first nationally and then between autonomous communities of Aragon, Basque Country and Andalusia using descriptive variables of the labor market. As Garcia (2017) points out, differences in labor indicators change with the economic cycle, and therefore a period of recession can influence the gender gap. It is observed that the differences in unemployment are reduced during the recession and that the size of the differences in wages differs if we talk about permanent contracts and temporary contracts. Finally, we constructed an econometric model involving 17 autonomous communities, except for Ceuta and Melilla, using the panel data technique. Consequently, this paper aims to analyze the evolution of the gender wage gap during the current crisis from a regional perspective, emphasizing the ability to explain the evolution of this gap in the difference in unemployment rates and in the productive structure of the regions.

ÍNDICE

1 Introducción	4
2 Análisis descriptivo general de las diferencias de género: el efecto de la crisis	5
3 Análisis descriptivo regional	13
4 Modelo econométrico: metodología y base de datos	22
5 Análisis de econométrico: resultados	31
6 Conclusiones	38
7 Bibliografía	39

1.- Introducción

El presente trabajo enfoca su base de estudio en el análisis del mercado laboral español centrado principalmente en el periodo de recesión tras el inicio de la crisis del año 2007 y analizando sus efectos sobre las diferencias de género utilizando series de datos recopilados de diferentes instituciones públicas como el Instituto Nacional de Estadística (INE) y el Ministerio de Educación, Cultura y Deporte hasta el año 2014. Estudiaremos la evolución de las diferencias salariales por género entre diferentes Comunidades Autónomas utilizando variables descriptivas del mercado laboral como es la tasa de actividad, la tasa de paro, la tasa de ocupación, el salario por hora normal trabajada, la evolución del sector industrial y de servicios respecto del PIB y los años medios de educación. Este tema es de actualidad, como se discute en Murillo y Simón (2014) o en DEL RÍO y VILLAR (2014) quienes han examinado la evolución de las diferencias salariales entre hombres y mujeres en España. Estas diferencias afectan a cuestiones tan relevantes como el riesgo de pobreza, y Cantó, Cebrián y Moreno, (2016) muestran que la incidencia que las mujeres entre 16 y 44 años tienen un mayor riesgo relativo de pobreza en 2013 que la que tenían antes de la crisis.

Como señala García (2017) las diferencias en los indicadores laborales cambian con el ciclo económico, y por tanto un periodo de recesión puede influir en la brecha por género. Desde una perspectiva agregada, De la Rica and Rebollo-Sanz (2017) and Peña-Boquete (2014) encuentran que la brecha en el desempleo disminuye en los periodos de recesión debido a que el empleo femenino se concentra en sectores menos sensibles al ciclo económico, Guner et al. (2014).

Por otro lado, como se indica en, Huertas et al. (2017), estas diferencias son heterogéneas desde el punto de vista regional, puesto que la estructura productiva de las CCAA difiere y esta estructura afecta a la brecha por género.

En consecuencia, este trabajo pretende analizar la evolución de la brecha salarial por género durante la actual crisis desde una perspectiva regional, poniendo el énfasis en la capacidad de explicar la evolución de dicha brecha de la diferencia en tasas de paro y en la estructura productiva de las regiones. No se pretende cuantificar la discriminación, sino simplemente ver cómo las condiciones económicas han influido en dicha brecha.

Comenzaremos con un análisis descriptivo general de las diferencias entre hombres y mujeres o viceversa en el mercado laboral de la economía española enfocándonos en la evolución de los datos de paro, ocupación, actividad y salarios.

A continuación, profundizamos el análisis y nos centramos en varias regiones situadas en distintos puntos de la península española. Estas regiones son Andalucía, Aragón y el País Vasco.

Seguidamente realizaremos un análisis econométrico involucrando a todas las Comunidades Autónomas españolas a excepción de Ceuta y Melilla. La técnica para analizar 17 comunidades autónomas elegidas será mediante datos de panel. Los datos de panel se forman de un grupo de individuos, en este caso 17 comunidades autónomas, observadas en un periodo de tiempo (2008-2014). Dentro de las ventajas de usar datos de panel y una de las más importantes, es poder encontrar una estructura más compleja y consecuentemente poder resolver problemas más difíciles que no se podrían solucionar usando solamente un enfoque puro de series temporales o de datos de sección cruzada.

Por último, nos preguntaremos si esas diferencias pueden ser explicadas por algunas teorías o modelos como el modelo de segregación ocupacional.

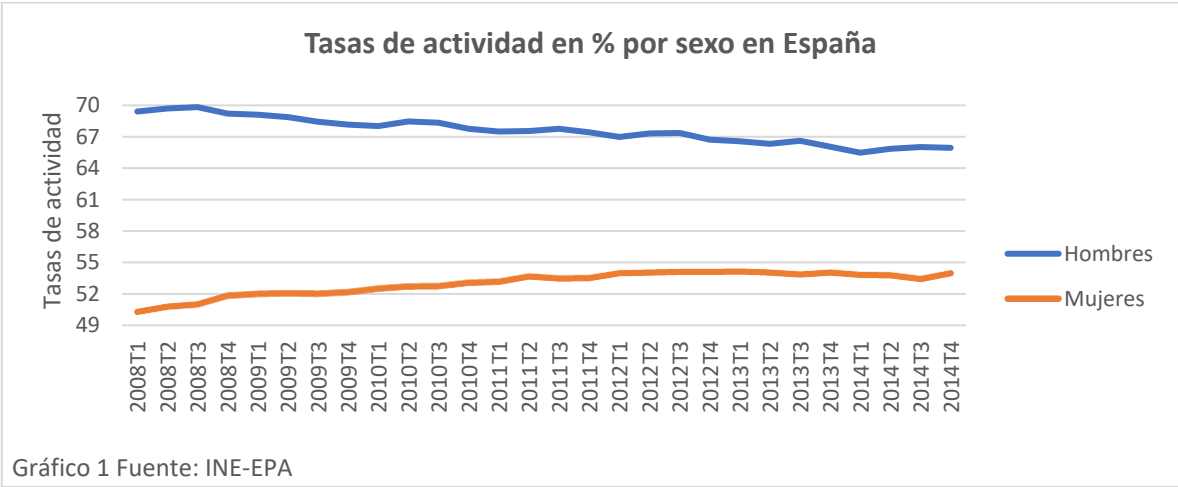
Para finalizar, extraeremos las conclusiones y las resumiremos en un breve apartado.

2.- Análisis descriptivo general de las diferencias de género: el efecto de la crisis

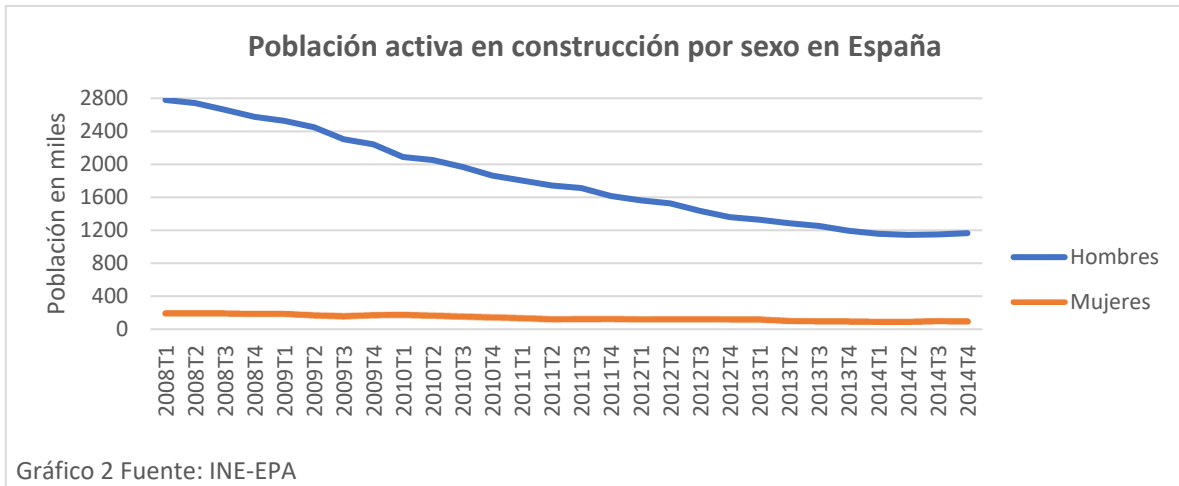
Desde el 2008 las economías occidentales han experimentado una profunda crisis económica. Esta crisis ha supuesto una importante transformación de las relaciones económicas, tanto en términos de estructura productiva, puesto que no ha afectado de manera homogénea a todos los sectores de actividad, como dentro de cada sector, dado que las reformas laborales auspiciadas por los gobiernos han modificado las condiciones en las que se realizan los contratos laborales.

En este sentido, es interesante analizar cómo la crisis ha afectado a las diferencias laborales por género que se veían observando en la economía española, tanto a nivel global como autonómica.

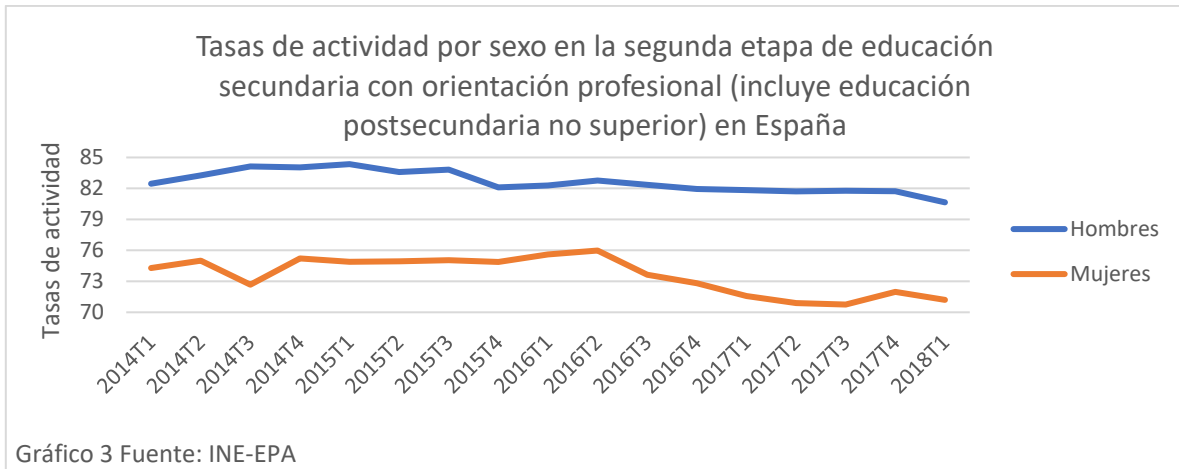
En la evolución de la tasa de actividad en el mercado laboral español durante el periodo de recesión posterior a 2007 hasta 2014 se puede observar una convergencia entre la tasa de actividad de los hombres con la de las mujeres. Esta convergencia viene marcada por el descenso de la tasa de actividad masculina (superior a don puntos) y no por el incremento de la femenina.



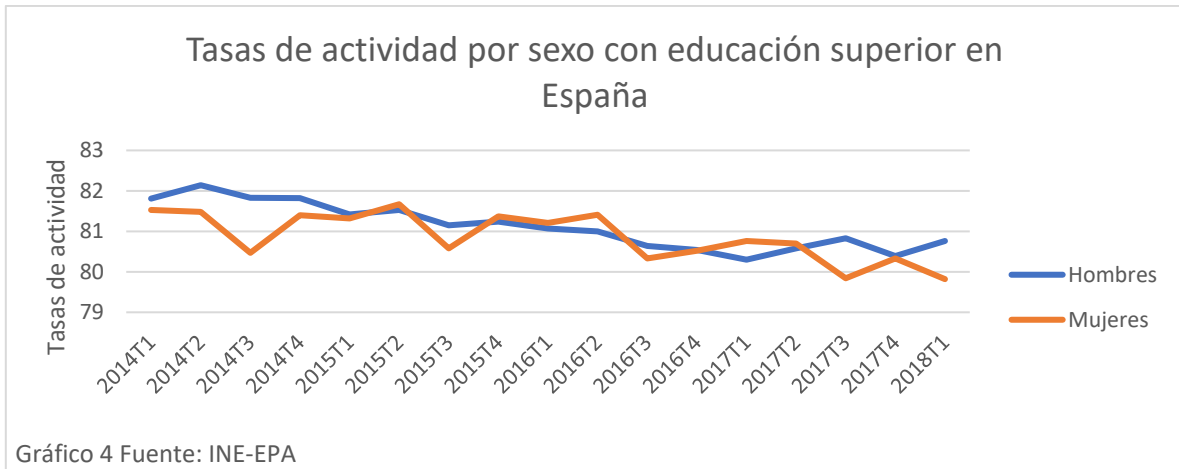
Este descenso de la tasa de actividad masculina se explica en buena parte por la gran sacudida que ha experimentado el sector de la construcción que se ha traducido en la pérdida de una inmensa cantidad de empleo, y marca la tendencia hacia el abandono del mercado de trabajo de los hombres.



Las diferencias de género en la tasa de actividad entre los niveles de educación primarios y secundarios son notables porque se alcanzan diferencias que rondan los 10 puntos porcentuales entre hombres y mujeres.

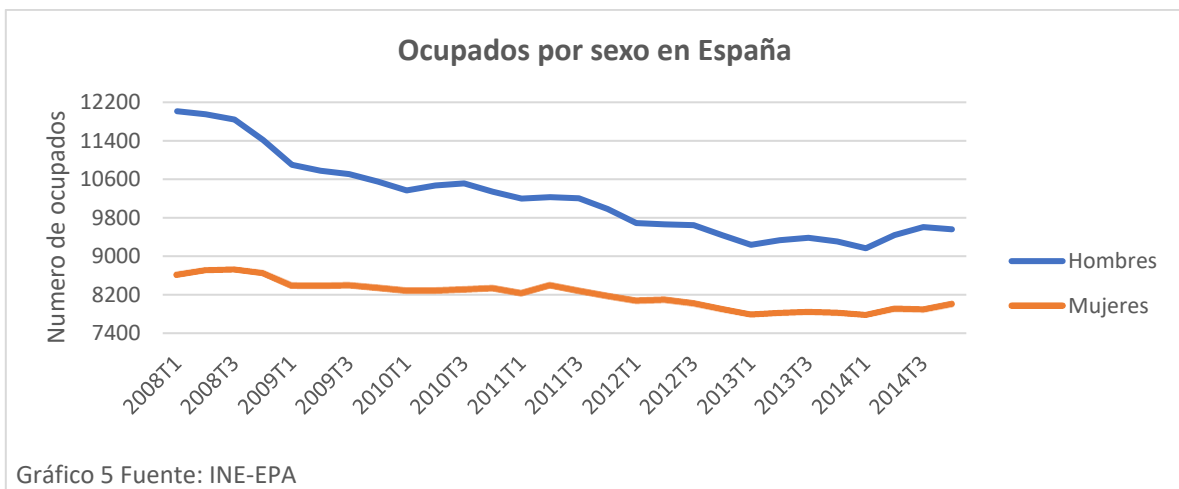


Filtrando las tasas de actividad por niveles de educación se puede observar que en la rama de educación superior las diferencias en tasas de actividad se invierten favoreciendo el género femenino y por tanto perjudicando al género masculino. Sin embargo, cabe mencionar que esas diferencias, en el periodo 2014-2018T1, fluctúan entre 0 y 2,5 puntos porcentuales. Un dato interesante que añadir es que los no casados/casadas tienen una tasa de actividad superior a los casados/casadas, pero unas diferencias de género superiores.

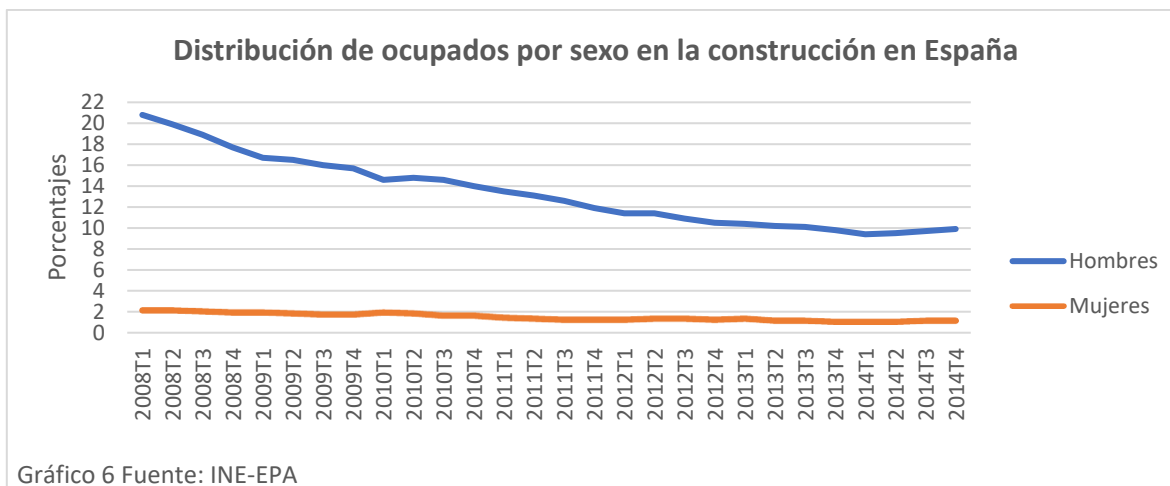


En contraste o en comparación a lo anterior podemos decir que la diferencia de tasas de actividad en educación superior es relativamente pequeña a la diferencia en niveles de educación inferiores al de educación superior.

El diferencial de género en el nivel de ocupación ha disminuido con el transcurso de la crisis a produciéndose así una convergencia relativa a razón de que el número de hombres ocupados ha disminuido en más de 2 millones de ocupados mientras que el número de mujeres ocupadas se ha mantenido estable alrededor de 8 millones.

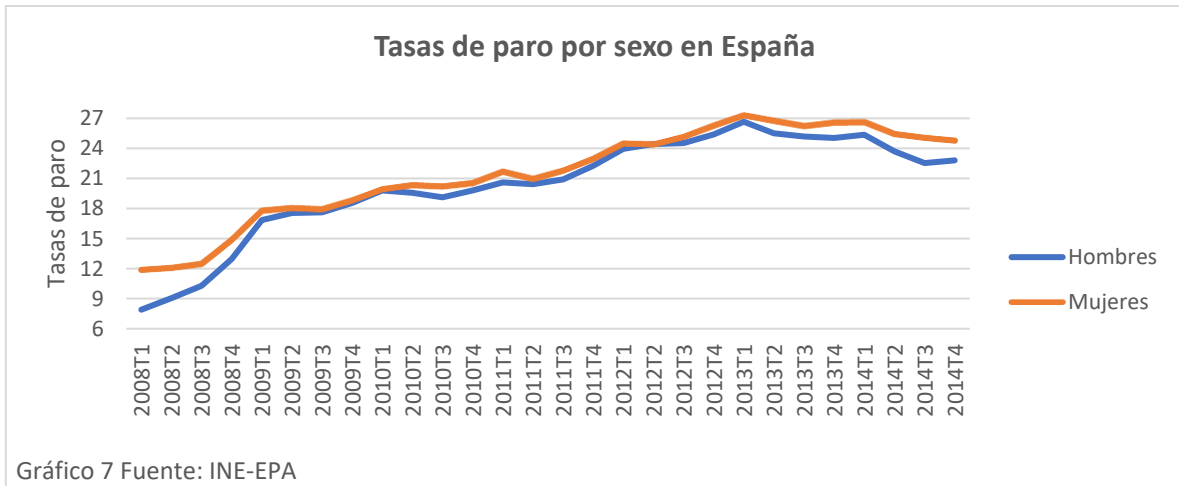


La variación negativa en el número de hombres ocupados es explicada en buena parte por el impacto de la crisis en el sector de la construcción el cual pasó de 21 millones de ocupados en el primer trimestre de 2008 a casi 11 millones a finales de 2014.

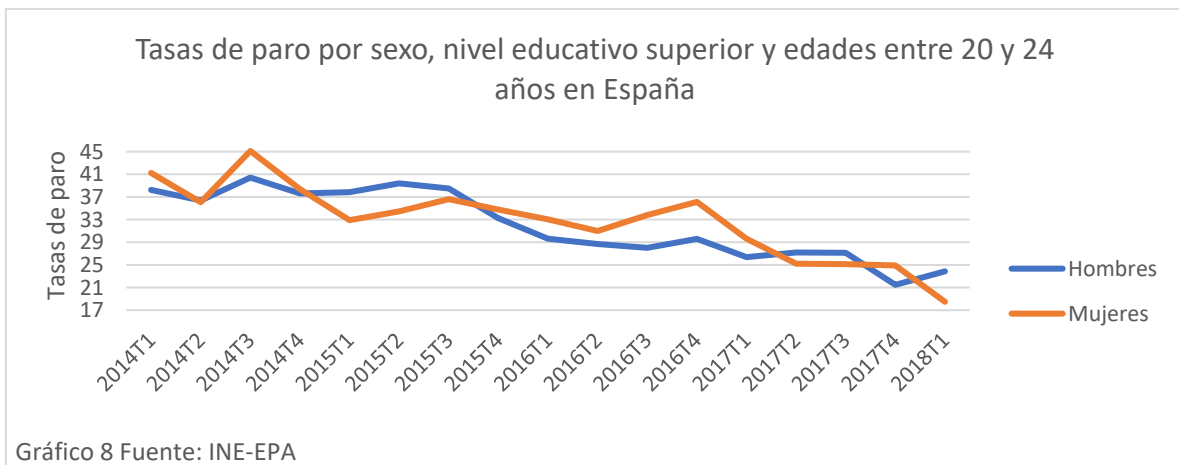


Todos estos hombres que perdieron el empleo parece que han migrado al sector servicios, claramente dominado por mujeres y por una gran diferencia todavía, aunque cabe apreciar que hubo una convergencia en las tasas de ocupación entre mujeres y hombres cuya tasa de ocupación pasó del 51% a principios de 2008 a 65% a finales de 2014.

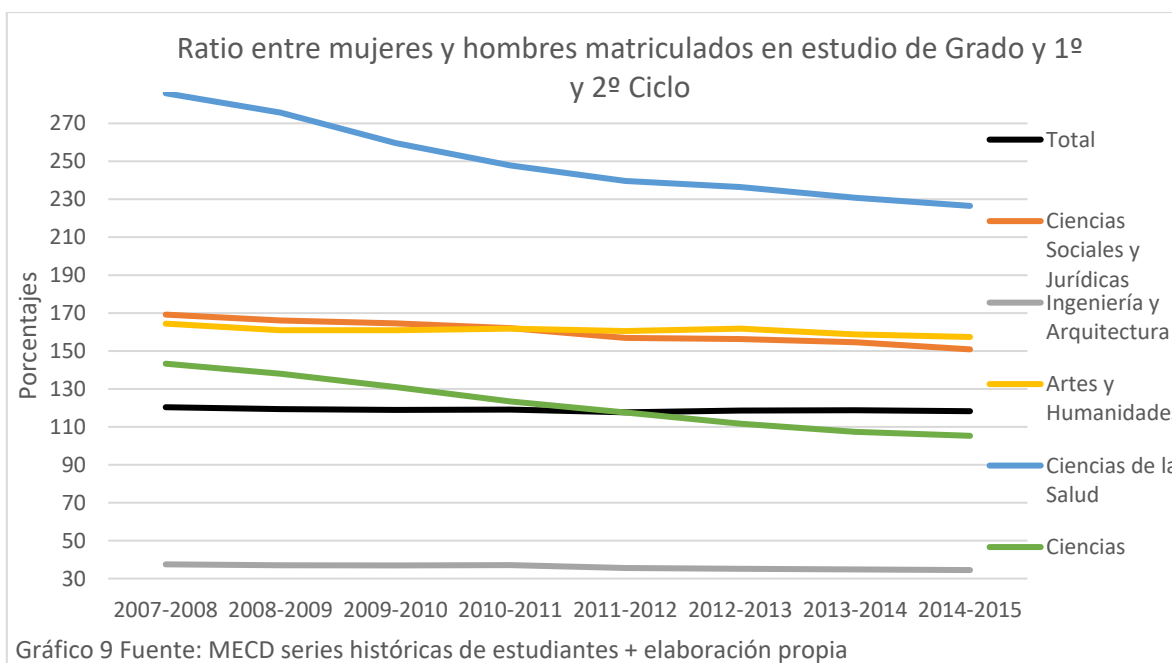
En términos nacionales podemos decir que no se observan diferencias significativas en las tasas de paro de hombres y mujeres durante el periodo de recesión en España. Aun así, comentaremos las pequeñas variaciones. De 2008 a 2009 se observa una convergencia de la tasa de paro masculina a la femenina incrementándose 10 puntos porcentuales en tan solo un año. A partir de este periodo ambas tasas se mantienen unidas hasta 2013, año a partir del cual la tasa de paro masculina decrece a una velocidad ligeramente mayor a lo que lo hace la tasa de paro femenina. Ello podría ser porque el sector de la construcción comienza a reactivarse lentamente combinado con la migración e integración de los hombres desde el sector de la construcción en el sector servicios.



Teniendo en cuenta la educación y según los datos del INE podemos decir que los adultos de mediana edad y maduros (29 años en adelante) presentan pequeñas diferencias en las tasas de paro relativamente favorecedoras con el hombre. Sin embargo, observamos que los adultos jóvenes (<29 años) parece que desde la parte final de la recesión las tasas se han convergido en una trayectoria decreciente pasando del 40% en 2014 al 22% a finales de 2017 para ambos géneros. Esto podría ser los efectos de las políticas de empoderamiento enfocadas al género femenino promovidas por el feminismo contemporáneo. Se observa una tendencia más decreciente de la tasa de paro de las mujeres que de los hombres. Podríamos fijarnos en el último dato, 2018T1, donde vemos una inversión de la situación. Si consideramos seguirán así en el futuro se podría ver como un síntoma de que las políticas educativas del pasado han sido sesgadas hacia las chicas por la presión de la tercera ola feminista y perjudicado a los chicos como se habla en “The war gainst Boys” de la autora Christina Hoff Sommers.



Otro indicio sería el creciente desequilibrio de los estudiantes y graduados universitarios. Según un documento del INE “ En España en el año 2015, el porcentaje de mujeres graduadas en educación superior era un 53,1% y el de hombres 46,9%.”. Por otro lado, según datos del Ministerio de Educación, Cultura y Deporte observamos la siguiente evolución de los datos de matriculados en estudios de Grado y 1º y 2º Ciclo de mujeres y hombres.



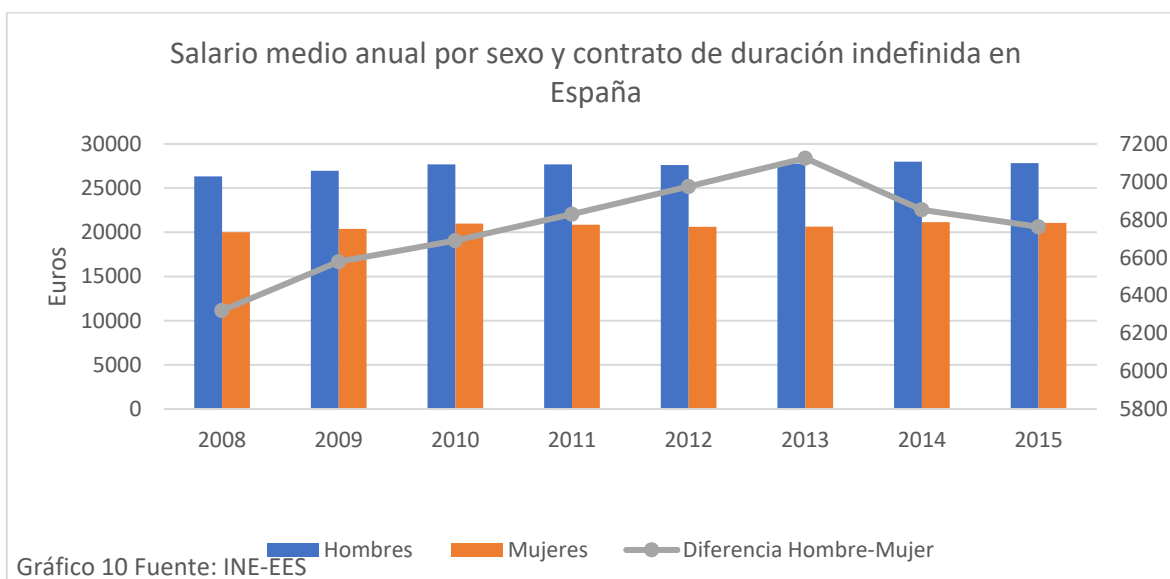
Comparando los matriculados en el curso 2008-2009 con los matriculados en el curso 2014-2015 podemos decir la diferencia total entre mujeres y hombres tiene una tendencia ligeramente decreciente durante este periodo de crisis. La ratio total mujer/hombre ha pasado de 120,35% en el curso 2008-2008 a 118,27% en el curso 2014-2015, es decir, tan solo un 2%. Aquí queda evidenciado que la educación está dominada por mujeres.

Desglosando el total por ramas de enseñanza destaca la participación de mujeres en Ciencias de la Salud donde en el curso 2007-2008 por cada 100 hombres matriculados, había 286 mujeres matriculadas y tras 7 años la cifra disminuyó a 230 mujeres por cada 100 hombres matriculados. Una muy notable desproporción de mujeres respecto de hombres también se da en las ramas de Ciencias Sociales y Jurídicas y Artes y Humanidades. La mayor similitud o disminución de la desigualdad entre el número de mujeres y hombres se

da en la rama de Ciencias donde había, en el curso 2007-2008, 143 mujeres por cada 100 hombres a 107 mujeres por cada 100 hombres en el curso 2014-2015.

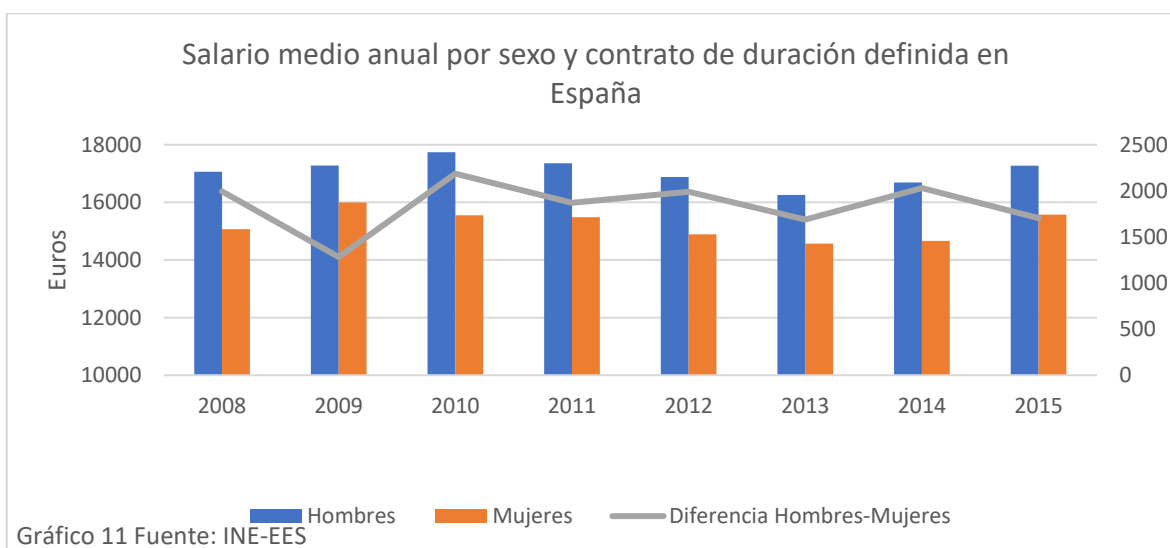
Por el contrario, la única rama dominada numéricamente por hombres es Ingeniería y Arquitectura. En el curso 2007-2008 hubo 37,51 mujeres matriculada por cada 100 hombres y en el curso 2014-2015 esa cifra decreció hasta 34,83 mujeres por cada 100 hombres.

Si analizamos las diferencias salariales entre hombres y mujeres en España podemos decir que, en general, los hombres con contrato de duración indefinida tienen mayores sueldos anuales que las mujeres y que esta diferencia ha crecido durante la crisis. Concretamente en 2008 la diferencia entre el salario anual indefinido de un hombre y una mujer era de 6320,52€ y ha incrementado hasta los 6852,33€ lo cual supone un incremento de 531,81€ de diferencia.



Por otro lado, las diferencias salariales medias anuales entre hombres y mujeres con contratos de duración determinada son mucho menores que los de duración indefinida pero también mucho más bajos. La diferencia comienza en 2008 con 1993,50€ y termina siendo 2029,67€ en 2014 lo cual supone un incremento de la diferencia de 36,17€. En 2015 la diferencia disminuye en 326,56€ pasando a una brecha de 1703,11€. A pesar de que las diferencias en la tasa de paro han disminuido, las diferencias salariales han aumentado. La

razón de ello reside en que las diferencias de paro merman porque incrementa el desempleo masculino, en concreto, se extinguen los contratos de los trabajos de baja cualificación y estos generalmente presentan abundancia masculina. Al reducirse los trabajos de baja cualificación y, por lógica, de menor retribución salarial se presenta una nueva situación en la que la media salarial de los hombres es ahora relativamente superior a la que había cuando en el mercado abundaban los trabajos con salarios más bajos y, por tanto, en media, las diferencias aumentan.



3.- Análisis descriptivo regional

Realizaremos un análisis de las diferencias entre hombres y mujeres en las tasas de actividad, ocupación, paro y unos índices salariales por comunidades autónomas. Elegimos 3 comunidades autónomas sobre las cuales aplicar el análisis. Primero una comunidad autónoma en general mejor posicionada que la media de España en términos de paro, empleo, actividad y salarios como es el caso de País Vasco. Segundo una comunidad autónoma en la otra parte del espectro, con abundante paro y menores salarios, Andalucía. Por último, la comunidad autónoma en la que estamos, Aragón, que está prácticamente en la media de España en muchos aspectos. El periodo temporal para el análisis será desde 2008 hasta 2014.

En términos de tasa de actividad, la primera observación que hacemos es que en términos nacionales las tasas de actividad de los hombres es 7 puntos porcentuales superior a la tasa de actividad nacional en el primer trimestre de 2008 mientras que la de las mujeres fue 10 puntos porcentuales inferior a la media. Cabe decir que durante el transcurso de la crisis esta diferencia entre tasa de actividad entre hombres y mujeres ha ido disminuyendo continuada y notoriamente con el transcurso de la crisis por dos motivos simultáneos. Por un lado, la tasa de actividad masculina siguió una tendencia decreciente y por otro lado la tasa de actividad femenina siguió una tendencia alcista.

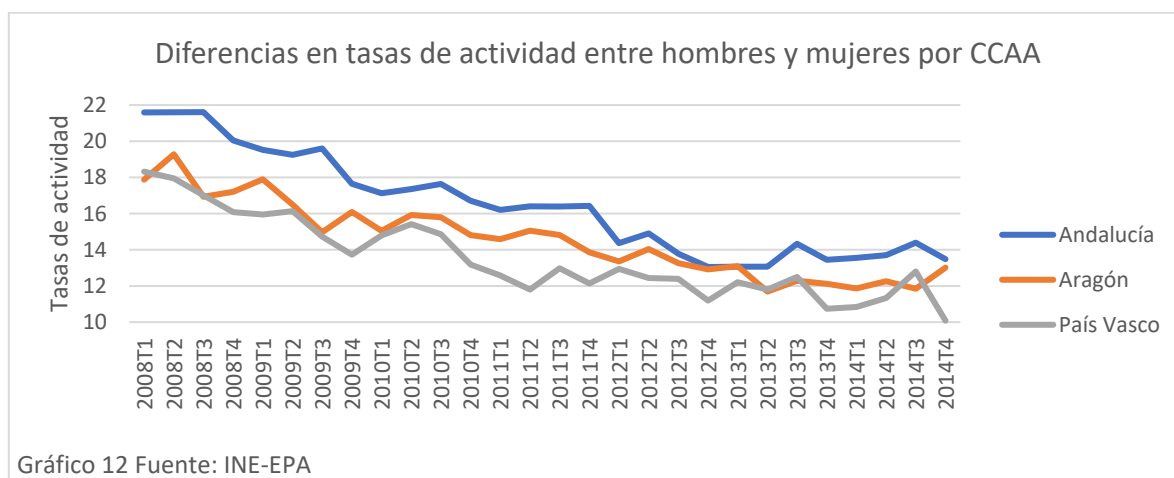
En Andalucía la tasa de actividad masculina sufre una disminución muy pequeña entre 2008 y 2014 de 2 puntos porcentuales (desde un 68% hasta un 66%) mientras que la tasa de actividad femenina experimenta una subida relativa mucho mayor, desde un 46% en 2008 hasta el 53% a finales de 2014 lo que supone un incremento de 7% porcentuales lo que supone una convergencia de las diferencias en las tasas.

La diferencia entre tasas de actividad de hombres y mujeres en Andalucía tiene una tendencia decreciente durante todo el periodo de recesión. Comienza la recesión con un diferencial del 22%, es decir, la tasa de actividad de los hombres es superior en 22 puntos porcentuales de la tasa de actividad de las mujeres. A finales del periodo recesivo el diferencial se sitúa en torno al 14%. Esto se debe a que la tasa de actividad de los hombres ha disminuido de 68% en 2008T1 a 66% en 2014T4 y la tasa de actividad de las mujeres se ha incrementado notablemente desde el 46% en 2008T1 a 52% en 2014T4.

Aragón por su parte ya presentaba de una desigualdad de género en las tasas de actividad relativamente inferior a Andalucía debido a que, pese a que ambas tasas de actividad masculinas estaban comprendidas entre 65% y 70%, la tasa de actividad femenina en Aragón ya partía de una posición superior del 50% comparado al 46% en Andalucía. Podemos observar una clara convergencia entre las tasas durante la crisis, pero llama la atención que, en Aragón, la tasa de actividad femenina parece estancarse en el periodo 2011-2014 donde se mantuvo en el 52% a diferencia de la tasa de actividad femenina en Andalucía que continuó creciendo y superando a la de Aragón. Ello se traduce en un

diferencial con tendencia decreciente, del 18% a comienzos del periodo recesivo al 13% en su parte final, e inferior al andaluz.

País Vasco parte de una situación en la que está mejor posicionada que Andalucía y Aragón. El diferencial sigue una tendencia decreciente desde un 18% al igual que Aragón e inferior que Andalucía y finaliza el periodo estando en el 10%. Ello supone una reducción de las diferencias del 8%, muy similar a la reducción en Andalucía. Una vez más, la convergencia entre tasas se da debido a la reducción de la tasa de actividad de los hombres de un 4% durante la recesión y un aumento de la tasa de actividad de las mujeres del 4% también.



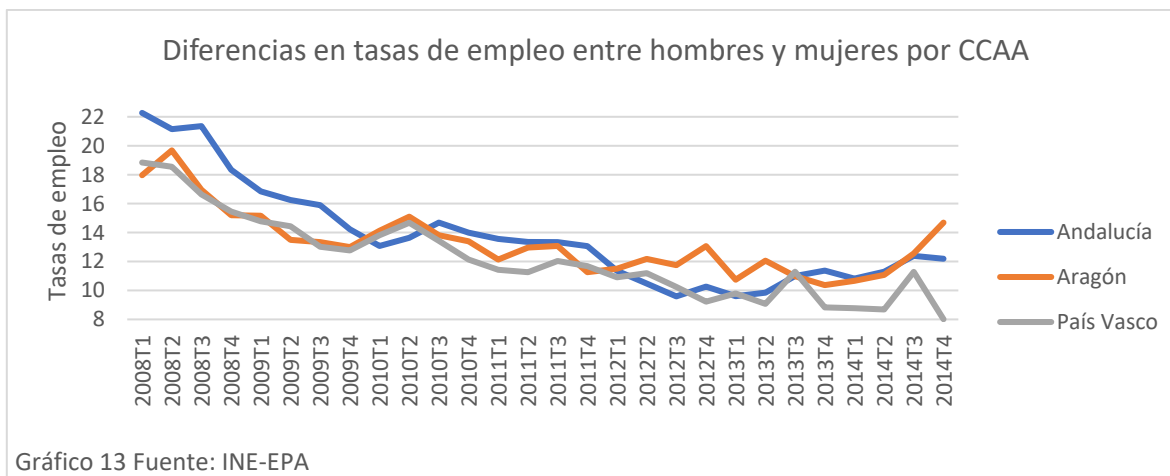
La comunidad autónoma que ha conseguido reducir más la desigualdad de hombres y mujeres fue País Vasco ya que pasó de una diferencia de 18,32% en el primer trimestre del año 2008 a 10,08% en el cuarto trimestre de 2014 lo que supuso una reducción de 8,61% seguido de 8,11% en Aragón y 7,12% en Andalucía.

En términos nacionales la tasa de empleo masculina es superior a la tasa de empleo nacional y la tasa de empleo de las mujeres inferior a la media nacional. Se observa que en 2008 la tasa de empleo masculina ha sufrido una caída muy grande a causa de, principalmente, el golpe al sector de la construcción. Desde 2009 y hasta terminar el periodo de recesión, ambas tasas de empleo han seguido unas variaciones prácticamente idénticas. Se observa una convergencia hasta el año 2011, año a partir del cual las diferencias de mantienen constantes.

En Andalucía ambas tasas tienen una tendencia decreciente pero la masculina presenta una pendiente más abrupta lo cual lleva a una convergencia entre las tasas de empleo de hombres y mujeres particularmente desde 2008 hasta 2012 ya que a partir de 2013 observamos que la tasa de empleo de hombres comienza una tendencia creciente mientras que la tasa de empleo femenino se mantiene constante por lo que, en este periodo, la desigualdad aumenta, aunque en términos generales se reduce. Ahora bien, en términos del diferencial entre tasa de empleo masculina y femenina se aprecia una abrupta tendencia decreciente desde comienzos del periodo recesivo (22%) hasta el año 2012T3 (10%), momento a partir del cual el diferencial comienza a incrementarse hasta el 12% a finales del periodo.

Aragón parte de unas tasas de empleo masculino y femenino superiores a las de Andalucía. La tasa de empleo femenino ha experimentado una caída más fuerte en Aragón que en Andalucía mientras que la caída en la tasa de empleo masculino en Aragón fue menor que en Andalucía lo cual sugiere pensar que el impacto en el sector de la construcción fue menos severo en Aragón. El diferencial presenta una tendencia negativa partiendo desde un 18% hasta un 12% en el último trimestre de 2011, momento a partir del cual el diferencial inicia un periodo de tendencia creciente hasta finales de 2014 situándose en casi un 15%. En términos netos la diferencia ha disminuido durante la crisis.

La tasa de ocupación en el País Vasco sigue un comportamiento similar al de Andalucía. La tasa de empleo masculino presenta una tendencia decreciente más abrupta que la tasa de empleo femenino. Puede decirse que las principales razones fueron el impacto en el sector de la construcción, donde la gran mayoría de los empleados son hombres, y el impacto en el sector financiero ya que País Vasco es un área donde el sector financiero es importante. Al contrario que Andalucía y Aragón el diferencial del País Vasco mantiene una tendencia negativa con mayor grado de consistencia. Comienza el periodo en 18% y lo finaliza en un 8% lo cual supera con creces a las otras dos comunidades autónomas.



Una vez más, País Vasco presenta una evolución más favorable hacia la minimización de diferencias en las tasas de empleo de hombres y mujeres que sus vecinos Andalucía y Aragón.

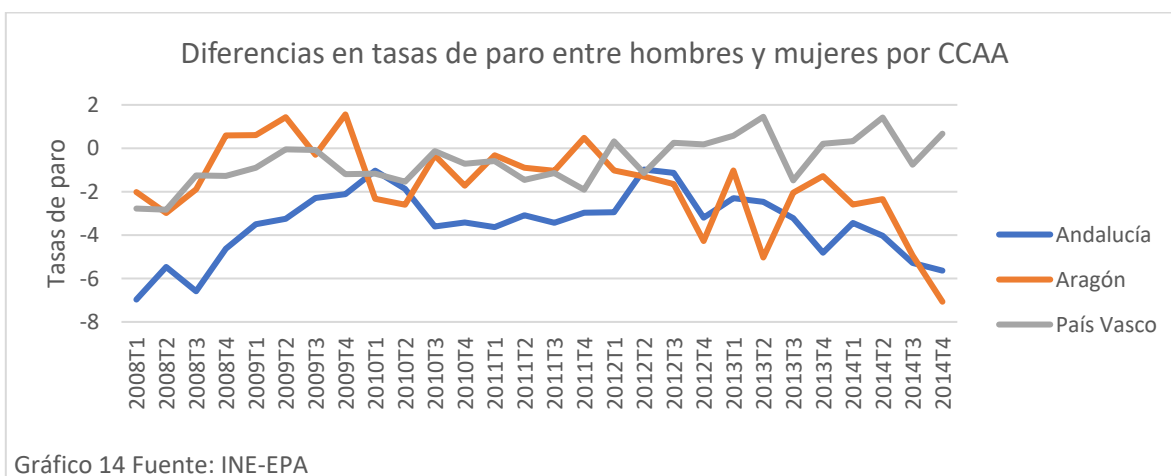
En términos nacionales las tasas de paro de hombres y mujeres experimentan un proceso de convergencia durante el año 2008 y, a partir de 2009, transcurren un periodo de igualdad hasta el año 2012. Tras finalizar éste se observa un pequeño incremento de desigualdad de tasas de paro entre hombres y mujeres en donde la de los hombres es ligeramente inferior a la de las mujeres.

Observando el diferencial de paro entre hombres y mujeres apreciamos una disminución de éste desde el -7% en 2008T1 hasta el -2% en 2010T1 probablemente porque la tasa de paro masculina aumenta debido a la caída del sector de la construcción. A partir de aquí se inicia un periodo donde este diferencial se mantiene en torno al -3% hasta comienzos del año 2012 momento a partir del cual el diferencial vuelve a aumentar hasta el -6% al final del periodo debido en parte a que el sector de la construcción comienza a mostrar signos de mejoras y a que esos hombres parados encuentran nuevos trabajos en distintos sectores del de la construcción.

La tasa de paro de Aragón está por debajo de la media española junto al País Vaco, al contrario de Andalucía. Las tasas de paro entre hombres y mujeres son muy similares y se van cruzando periodo tras periodo hasta mediados del año 2012. A partir de éste se percibe una divergencia entre las tasas, siendo mayor la tasa de desempleo de las mujeres ello

quizás a razón de que el sector de la construcción se esté reactivando. Observando el diferencial vemos que tiene una tendencia neta negativa, por lo que en Aragón podemos concluir que la crisis ha incrementado la diferencia en tasa de paro entre hombres y mujeres.

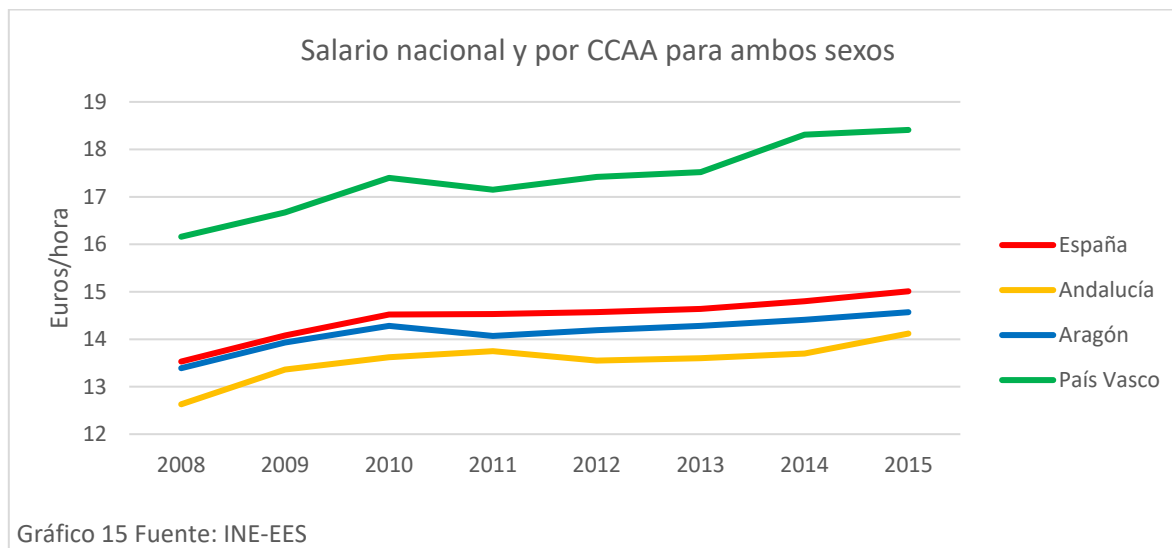
País Vasco presenta las menores tasas de paro de hombres y mujeres. Se percibe una convergencia total durante el año 2008 entre las tasas de desempleo masculino y femenino. Estas tasas presentan, a partir del año 2009 y hasta finales del periodo de recesión, una entrelazada evolución con tendencia positiva que nos permite decir que durante la crisis se ha contribuido a eliminar la desigualdad en las tasas de paro, aunque ya presentaban unas diferencias muy pequeñas de antemano. Fijándonos en el diferencial vemos que su tendencia positiva ha llevado a invertir las diferencias entre hombres y mujeres en cuanto a tasas de paro se refiere, es decir, al final del periodo la tasa de paro de las mujeres era inferior a la de los hombres.



En cuanto a salarios:

En términos generales observamos que los salarios por hora en País Vasco son muy superiores a la media española que ronda los 14,30€/hora. El salario por hora en Aragón se sitúa ligeramente por debajo de la media española pero alrededor de 50 céntimos por hora por encima del salario/hora andaluz. El salario hora aragonés y andaluz presentan un incremento 89 céntimos y 99 céntimos euro respectivamente mientras que el salario hora vasco se incrementó en 1,24€/hora. A partir de este periodo el salario hora de Aragón y

Andalucía se mantiene en una tendencia creciente con una pendiente muy próxima a cero mientras que el salario del País vasco continúa creciendo notablemente alcanzando los 18,31€/hora en 2014.

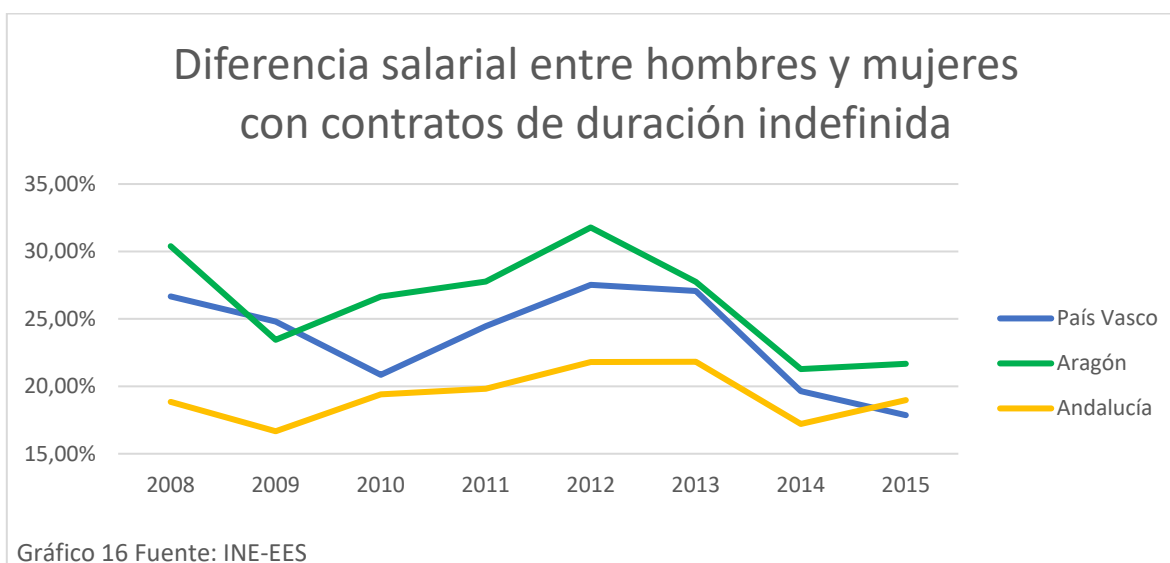


En Andalucía se observa que la diferencia salarial entre hombres y mujeres con contratos de duración indefinida entre 2008 y 2014 es menor que la diferencia en Aragón y País Vasco. Esta comienza con el primer año de recesión situándose en un 19% tras el cual desciende al 17% en 2009. A partir de este periodo comienza una tendencia alcista de la diferencia hasta 2013 superando el 21% año tras el cual experimenta una brusca caída hasta el 17%. En términos netos, entre 2008 y 2014, la diferencia en salarios entre hombres y mujeres ha tenido una tendencia ligeramente creciente, aunque si nos fijamos exclusivamente en los datos de 2008 y lo comparamos a los de 2014, la diferencia se redujo un 1,64%.

Aragón es la comunidad autónoma que tiene que mayores diferencias, pero también es la que mayor reducción de esta experimenta. Del 2008 al 2009 sigue el mismo patrón que Andalucía, pero de forma más agresiva puesto que la diferencia decrece del 30% al 23% en tan sólo un año respectivamente. A continuación, inicia una etapa de crecimiento de las diferencias que alcanza su cúspide en 2012 con casi un 32% y en los dos años posteriores experimenta una tendencia inversa a la anterior con la misma intensidad hasta llegar al 21% en 2014. Se ve claramente que las diferencias se reducen en términos generales en el

periodo final de la crisis mientras que aumenta en el periodo intermedio. En términos netos las diferencias disminuyen un 9,11% entre 2008 y 2014.

El País Vasco se sitúa en el intermedio de las otras comunidades autónomas. Sigue un patrón prácticamente idéntico al de Aragón, pero partiendo de una posición inferior y finalizando en una posición inferior también. Comienza en 2008 con una diferencia del 27% y decrece al 21% en 2010. A continuación, crece y decrece agresivamente, muy similar a Aragón, hasta ser la diferencia menor a 20% en 2014. Comparando los años 2008 y 2014 la diferencia ha disminuido un 7,03%.

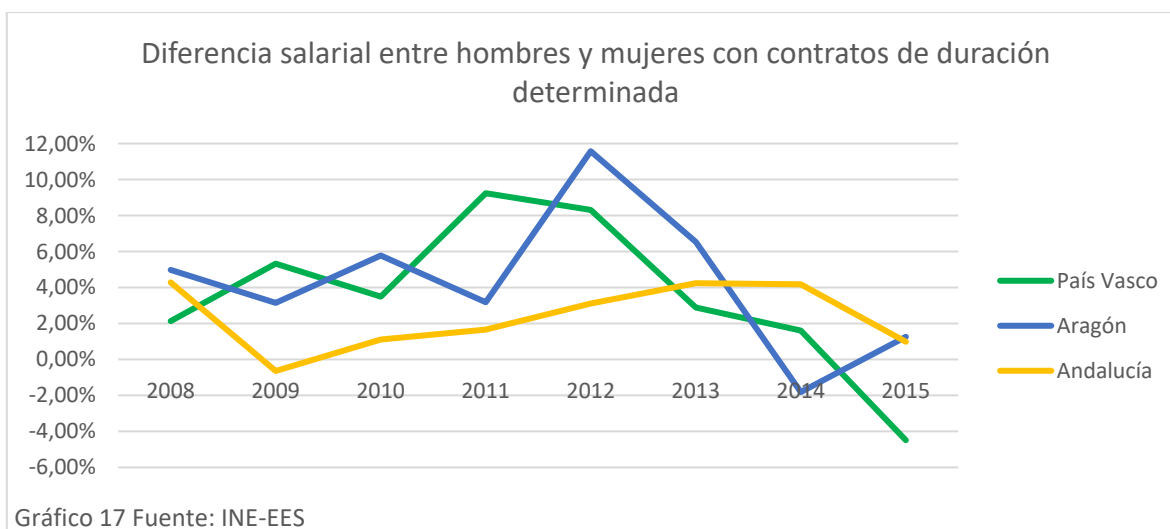


Llama la atención que las diferencias de salarios por hora entre hombres y mujeres con contrato de duración determinada son un tercio o menos de las diferencias entre hombres y mujeres contratos de duración indefinida.

Andalucía, entre el año 2008 y 2009 de da una disminución de la diferencia desde el 4,28% hasta un -0,62% respectivamente, es decir, en 2009 los hombres con contrato de duración determinada cobraban un 0,62% menos por hora que las mujeres con el mismo tipo de contrato. A partir de 2009 la diferencia crece hasta que en 2014 llega a los mismos valores que en el inicio de la crisis, un 4%. En términos netos, desde 2008 hasta 2014 se redujo la diferencia en 10 décimas porcentuales.

En Aragón la diferencia fluye alrededor del 5% hasta 2011, año después del cual experimenta una subida brutal hasta casi el 12% en 2012 y seguidamente baja con la misma intensidad hasta 2014, invirtiéndose la brecha salarial hasta el -1,81%, es decir, los hombres con contrato de duración determinada cobran un 1,81% menos de las mujeres con el mismo contrato. Se concluye que la brecha de salarios ha disminuido durante la recesión.

País Vasco experimenta una evolución similar a la de Aragón. En los comienzos del periodo recesivo la diferencia crece muy poco y durante el periodo central (que es cuando España experimenta la mayor caída del PIB) incrementa hasta doblarse para después caer en picado. En 2008 la diferencia fue de 2,13% y en 2014 del 1,60% lo que supone una reducción de 0,53%. La brecha se invierte después de la crisis, siendo los hombres los que cobran un 4,29% menos que las mujeres.



¿Por qué las diferencias crecen en el corazón de la crisis? Quizás sea porque se eliminan de las operaciones los salarios procedentes de trabajos menos productivos y por tanto menores. Estos tienden a estar frecuentados más por hombres que por mujeres y por tanto al ser los hombres los que han sufrido más la crisis y por tanto se eliminan los salarios más bajos de las cuentas, esto nos da una media salarial relativamente mayor que la anterior para los hombres y por tanto crecen las diferencias.

¿Por qué las diferencias son menores entre hombres y mujeres con contrato de duración determinada que de duración indefinida? Podría ser, en parte, que los trabajadores con contratos indefinidos consiguen mantener los beneficios negociados en el pasado y los

contratos de duración indefinida finalizan y se vuelven a crear con unas reglas más actualizadas y con mayor flexibilidad y con una mayor orientación hacia la igualdad.

4.- Análisis econométrico: metodología y base de datos

Las variables seleccionadas para explicar las diferencias salariales entre sexos son la tasa de actividad, la tasa de ocupación, la tasa de paro, y el salario por hora normal trabajada. Cada una de las variables previamente mencionadas vienen desglosadas por sexos. Son datos anuales y el periodo temporal se extiende desde 2008 hasta 2014. Además, el porcentaje que tiene el valor de lo producido sobre el PIB, por un lado, por el sector industrial y, por otro lado, el sector servicios. La fuente de estos datos se encuentra en el Instituto Nacional de Estadística (INE) y, en concreto, en la categoría de mercado laboral.

La técnica para analizar los datos de las 17 comunidades autónomas elegidas será mediante modelo de datos de panel. Los datos de panel se forman de un grupo de individuos, en este caso las 17 comunidades autónomas, observadas en un período de tiempo (2008-2014) (Carter Hill, Griffiths, & Lim, 2012). Es decir, los datos de panel¹ combinan cortes transversales (información de varios individuos en un momento dado) durante varios períodos de tiempo. El disponer de datos de panel constituye una ventaja y un inconveniente:

- Ventaja porque disponemos de más datos y se puede hacer un seguimiento de cada individuo.
- Inconveniente porque si todas las cualidades relevantes del individuo NO son observables entonces los errores individuales estarán correlacionados con las observaciones y los MCO serán inconsistentes.

En general, los datos se observan a intervalos regulares de tiempo. Los datos de panel pueden ser balanceados ($T_i = T$ para todo i) o no balanceados ($T_i \neq T$ para algún i). La

¹ Suponemos un panel balanceado (con todos sus datos completos). Un panel no balanceado es un panel en el que faltan algunas observaciones que se excluyen del cálculo. En este caso el sesgo también puede venir dado por la calidad de las variables observadas y la razón de que se omitan algunas observaciones

selección muestral debe ser aleatoria (no correlacionada con los regresores) para que los estimadores sean consistentes.

Se pueden tener paneles:

- de muchos individuos y pocos periodos temporales (“short panels”)
- de pocos individuos y muchos periodos temporales (“long panels”)
- de muchos individuos y muchos periodos temporales

Supongamos que el modelo que pretendemos estimar es el siguiente:

$$y_{it} = \alpha + \beta_1 X_{1it} + \dots + \beta_k X_{kit} + u_{it}$$

$i = 1, 2, \dots, N$
 $t = 1, \dots, T$
 $k = 1, \dots, K$

(1)

donde

- X_{1it}, \dots, X_{kit} son las k variables explicativas (observables). El subíndice i hace referencia al individuo, empresa, etc., el subíndice t indica el momento del tiempo.
- $\beta_1, \beta_2, \dots, \beta_k$ son los k parámetros a estimar y que miden el efecto de cada variable explicativa sobre la variable dependiente.
- u_{it} es el término de error que debe cumplir los supuestos básico de un modelo econométrico: esperanza cero, varianza constante (homocedasticidad), no autocorrelación y distribución normal (ruido blanco o i.i.d)

Es decir, se busca determinar si las variaciones observadas en y se deben a cambios en las variables explicativas, tomando en cuenta las diferencias individuales.

En el modelo tradicional de regresión lineal (Modelo Lineal General (MLG)) los parámetros a estimar se supone que son **invariantes en el tiempo y espacio**. La perturbación aleatoria del modelo recoge las diferencias entre individuos (efecto individuo) y a través del tiempo. Este modelo ignora la naturaleza de cada unidad transversal (empresa) y su evolución en el tiempo \Rightarrow **errores de especificación**. Los efectos individuales son aquellos que afectan de manera desigual a cada uno de los agentes de estudio contenidos en la muestra (individuos, empresas, bancos). Son invariables en el tiempo y afectan de manera directa las decisiones

que tomen dichas unidades (capacidad empresarial, eficiencia operativa, experiencia, tecnología)

Existen diversas fuentes de variabilidad:

- **efecto individuo:** generalmente invariante en el tiempo. Representa el impacto directo de todas las características individuales **no observables** e invariantes en el tiempo sobre y_{it} .
- **efecto tiempo:** que puede asumirse invariante entre individuos; cada periodo tiene efectos específicos no observables.
- **efecto individuo-tiempo:** efectos cambiantes que pueden ser tanto determinísticos como estocásticos

Suponer que los coeficientes son iguales para los N individuos y/o los T instantes de tiempo puede resultar muy restrictivo. Por su parte, el caso extremo, donde se asume **no** **kit** variantes entre individuos y en el tiempo puede ser imposible de manejar. Se requiere por tanto introducir algunos supuestos (Greene, 2007; Wooldridge, 2002):

- **Supuesto 1: Exogeneidad contemporánea:** x_t y u_t ortogonales en el sentido de la media condicional (este supuesto restringe la relación para el mismo periodo de tiempo). $E[u_t | x_t] = 0$, para $t = 1, 2, \dots, T$.
- **Supuesto 2: Exogeneidad estricta:** extiende la restricción de ortogonalidad en la relación en cualquier periodo. $E[u_t | x_1, \dots, x_t] = 0$, para $t = 1, 2, \dots, T$.

El supuesto 1 puede resultar difícil de mantener cuando existen variables omitidas. Ello constituye la principal motivación para utilizar datos panel. Así (1) se transforma en

$$y_{it} = \beta_0 + \beta_1 X_{1it} + \dots + \beta_k X_{kit} + \underbrace{\alpha_i + \varepsilon_{it}}_{u_{it}} \quad (2)$$

$u_{it} = \alpha_i + \varepsilon_{it}$: término de error compuesto (inobservado) donde α_i : efectos individuales (heterogeneidad inobservada permanente en el tiempo) y ε_{it} : error idiosincrásico. La variable α_i captura todos los factores no observados que afectan a Y_{it} y que no cambian en el tiempo: por lo tanto no lleva subíndice t. A este modelo se le suele llamar modelo de efectos no observados o modelo de efectos fijos. El término de error que ya conocemos es

ε_{it} . Incluye factores no observados que afectan a Y_{it} y que cambian en el tiempo. Al término $u_{it} = \alpha_i + \varepsilon_{it}$ se le suele llamar error compuesto: tiene un componente fijo y otro que cambia en el tiempo

Para obtener estimaciones consistentes de los parámetros de posición (β_i) usando MCO debemos suponer que X_{it} no están correlacionada con u_{it} . Incluso si suponemos que $cov(X_{it}, u_{it}) = 0$, la estimación MCO será sesgada e inconsistente si $cov(X_{it}, \alpha_i) \neq 0$.

Existen dos modelos sustancialmente diferentes según el tratamiento de α_i :

Modelo de efectos fijos

Modelo de efectos aleatorios

Regresión agrupada (pooled)

Un modelo lineal estático para datos de panel:

$$\begin{aligned} y_{it} &= \beta_0 + \beta_1 X_{1it} + \dots + \beta_k X_{kit} + u_{it} \\ i &= 1, 2, \dots, N \\ t &= 1, \dots, T \end{aligned} \quad (3)$$

En el modelo dado en (3) todos los coeficientes se asumen constantes en el tiempo y a través de los individuos, de modo que el término de error recoja la heterogeneidad no observada. Se puede estimar consistentemente por MCO si se supone que los regresores son exógenos:

$$E[u_{it} | x_{1it}, \dots, x_{kit}] = 0$$

Sin embargo, los errores u_{it} del modelo no serán idénticos e idénticamente distribuidos (i.i.d). Es decir, las observaciones están agrupadas de forma natural por individuos i (“clusters”) probablemente existirá heterocedasticidad entre “clusters”. Para ello, sería necesario usar *los errores estándar robustos*, al menos por la presencia de “clusters”

El estimador MCO a partir de la regresión agrupada es simple y aprovecha tanto la variabilidad temporal como entre individuos de los datos. Como se ha mencionado, es posible que $Cov(X_{it}; u_{it}) \neq 0$, entonces los estimadores MCO a partir de la regresión

agrupada serán sesgados. Muchas veces dicha correlación es debida a un error de especificación por la ausencia de alguna variable relevante o la existencia de cualidades inobservables de cada individuo. Este problema puede solucionarse con una regresión de datos anidados.

Modelo de panel con Efectos fijos (MEF)

No debe olvidarse que el principal objetivo de los modelos de datos panel es precisamente capturar la heterogeneidad no observable y que es ignorada en los tradicionales modelos de regresión y que puede de alguna manera afectar la estimación de los efectos de las variables X sobre y. El MEF asume que las diferencias entre los individuos pueden ser capturadas a través de diferencias en el término constante, lo que equivale a asumir estas variaciones como determinísticas. Así, efecto fijo significa $cov(X_{it}, \alpha_i) \neq 0$. Siendo que se trata de variables no observadas, la heterogeneidad individual se recoge a través de un conjunto de N-1 variables dicotómicas (di), cuyos coeficientes asociados α_i indican las diferencias individuales con respecto al individuo de referencia y se estiman conjuntamente con las pendientes β_i .

La manera de hacerlo es incluir una variable binaria para cada unidad i. Cada variable binaria se denomina efecto fijo. Notar que no se trata de parámetros fijos, sino que simplemente para cada individuo son constantes en el periodo analizado.

Este modelo es el que implica menos suposiciones sobre el comportamiento de la perturbación aleatoria. Supone que el modelo a estimar es ahora:

$$y_{it} = a_i + \beta_1 X_{1it} + \dots + \beta_k X_{kit} + \varepsilon_{it} \quad (4)$$

Donde $a_i = \beta_0 + \alpha_i$, luego reemplazando en (4) queda:

$$y_{it} = \beta_0 + \beta_1 X_{1it} + \dots + \beta_k X_{kit} + \alpha_i + \varepsilon_{it} \quad (5)$$

Es decir, supone que el error (u_{it}) puede descomponerse en dos una parte fija, constante para cada individuo (α_i) y otra aleatoria que cumple los requisitos MCO (ε_{it}) → (

$u_{it} = \alpha_i + \varepsilon_{it}$), lo que es equivalente a obtener una tendencia general por regresión dando a

cada individuo un punto de origen (ordenadas) distinto. Esta operación puede realizarse de varias formas, una de ellas es introduciendo una dummy por cada individuo (eliminando una de ellas para evitar problemas de multicolinealidad exacta conocida como la trampa de variables ficticias) y estimando por MCO.

Otra es calculando las diferencias. Así, si (5) es cierto, también es cierto que:

$$\bar{y}_i = \lambda + \beta_1 \bar{X}_{1i} + \dots + \beta_k \bar{X}_{ki} + \alpha_i + \bar{\varepsilon}_i \quad (6)$$

Donde:

$$\bar{X}_i = \frac{\sum_{t=1}^{T_i} X_{it}}{T_i} \text{ es la media temporal para cada individuo de la muestra}$$

Es decir, transformar el modelo restando a cada variable su media individual:

$$(y_{it} - \bar{y}_i) = \beta_1 (X_{1it} - \bar{X}_{1i}) + \dots + \beta_k (X_{kit} - \bar{X}_{ki}) + (\varepsilon_{it} - \bar{\varepsilon}_i) \quad (7)$$

Este modelo se puede estimar consistentemente por MCO porque los regresores X_{1it} eran endógenos por su correlación con α_i , pero están incorrelados con ε_{it} (en cualquier periodo temporal). Cuando se disponga de estimaciones de β , se pueden obtener estimaciones de los efectos individuales:

$$\hat{\alpha}_i = \bar{y}_i - \hat{\beta}_1 \bar{X}_{1i} - \dots - \hat{\beta}_k \bar{X}_{ki} \quad (8)$$

Se deben utilizar errores estándar robustos si se piensa que ε_{it} no son i.i.d

La ventaja de usar estimadores de efectos fijos es que los coeficientes estimados son insesgados de manera que su sesgo es nulo por lo que la esperanza matemática es igual al parámetro que se desea estimar. Por otra parte una de sus desventajas es que esta estimación por efectos fijos no puede ser usada para investigar sucesos que no varíen en el tiempo en las variables dependientes.

Desventajas: al trabajar las variables en diferencia, se elimina el efecto de largo plazo. No puede estimarse el efecto de variables que no varíen en el tiempo tales como raza, sexo, educación, afiliación a sindicatos, religión, etc.

Modelo de Panel con Efectos aleatorios (MEA)

Hasta ahora hemos intentado controlar o eliminar α_i , pues se consideraba que α_i estaba correlacionada con x_{it} . Si la heterogeneidad individual no observada se asume no correlacionada con las variables incluidas en la matriz x ($\text{Cov}(X_{it}, \alpha_i) = 0$), el efecto individual constante puede considerarse se distribuye aleatoriamente entre los individuos del corte transversal. En ese caso se tiene un Modelo de Efecto Aleatorio (MEA), que puede expresarse como:

$$y_{it} = \beta_0 + \beta_1 X_{1it} + \dots + \beta_k X_{kit} + \mu_{it} \quad (9)$$
$$\mu_{it} = \alpha_i + \varepsilon_{it}$$

Donde α_i es el componente individual aleatorio similar a ε_{it} pero invariante en el tiempo (within), mientras ε_{it} es ruido blanco. Así, en el MEA se supone que α_i es una variable aleatoria inobservable independiente de x_{it} y que por tanto forma parte del término de perturbación compuesto, por lo que a este tipo de modelos también se les conoce como modelos de error compuesto. El término de error puede incluir también un componente temporal aleatorio, invariable entre individuos (between) $\mu_{it} = \alpha_i + \phi_t + \varepsilon_{it}$; sin embargo, por lo general se asume $\phi_t = 0$. Este tipo de modelos es adecuado cuando se trabaja con muestras muy grandes, extraídas de una población suficientemente grande, donde asumir interceptos diferentes puede resultar muy complejo. Al igual que en caso anterior, la estimación se hace a través de MCG (Baltagi, 1995).

Este modelo es más eficiente (la varianza de la estimación es menor) pero menos consistente que el de efectos fijos, es decir es más exacto en el cálculo del valor del parámetro pero este puede estar más sesgado que el de efectos fijos.

En efectos aleatorios tenemos que especificar las características individuales que pueden o no influir en las variables explicativas. El problema con esto es que algunas de las variables pueden no estar disponibles, por tanto, conduce a un sesgo de variables omitidas en el modelo (Torres-Reyna, 2007)

¿Qué significa que α_i es una variable aleatoria? Significa que no estamos seguros del valor exacto en el origen que pueda tener cada individuo sino que pensamos que este,

probablemente gravitará en torno a un valor central. Eso suele ocurrir cuando tomamos una muestra de un gran universo de individuos.

La diferencia con el modelo de efectos fijos frente al modelo de efectos aleatorios, es que en este último, el coeficiente para cada individuo se asume que es común para cada individuo y a lo largo del tiempo, más una variable aleatoria, α_i que varía entre individuos pero es constante dinámicamente, es decir, mide la variación aleatoria de cada coeficiente del término *global* " β_o ". El término α_i por otra parte tiene las características de un ruido blanco.

Contrastes de especificación

Surgen entonces dos dudas: ¿Cuándo debemos aplicar un MCO Pooled y cuando un modelo de datos de panel y, en este último caso, de entre los dos posibles cual de ambos es más procedente? Para solucionarlas debemos responder a varias preguntas:

- a) ¿la varianza de α_i es significativamente distinta de cero? Si la respuesta es afirmativa implica que efectivamente existe un componente inobservable de la varianza asociada a cada individuo y que MCO estará sesgado. Es decir, el contraste de modelo de datos de panel versus regresión agrupada (pooled) consiste en estimar si cada individuo tiene un origen en ordenadas distinto mediante la estimación de si α_i tiene una distribución distinta de cero. Nótese que tanto en el caso de efectos fijos (donde α_i tiene un valor constante para cada individuo pero una distribución para toda la muestra) como en el caso de efectos variables (donde α_i tiene una distribución para cada individuo) α_i siempre tiene que tener una cierta distribución (un valor y una desviación). Ojo, lo importante no es que tenga un valor, ya que el valor fijo se estima en la constante del modelo, sino que lo relevante es que tenga una varianza, una distribución, significativamente distinta de cero.
- b) Si la respuesta anterior es afirmativa, la siguiente es ¿Tenemos un panel en el que están TODOS los individuos del universo? En caso afirmativo se tienen que aplicar efectos fijos, si, por el contrario tenemos una muestra, más o menos representativa tendremos que pasar a la siguiente cuestión. El panel en el que están todos los

individuos del universo (por ej.: todas las provincias del país. todas las empresas de conservas del mercado, etc.) también se suele llamar (de forma no muy correcta) panel macro.

- c) Si la respuesta anterior es negativa, la siguiente pregunta es ¿las estimaciones consistentes (efectos fijos) y las eficientes (efectos aleatorios) son significativamente distintas? Una respuesta afirmativa implica que es mejor escoger el estimador que consideramos más consistente (el de efectos fijos), por el contrario si son ortogonalmente iguales se deberá escoger la estimación más eficiente, la de efectos aleatorios.

- A la primera pregunta se puede responder mediante el uso del contraste de Breusch-Pagan, también denominado del Multiplicador de Lagrange. La prueba consiste en realizar la regresión auxiliar $indepit = depit + u_i + eit$. La hipótesis nula es:

$$H_0 : \text{var}(u_{it}) = 0 \Leftrightarrow H_0 : \text{regresion agrupada (pooled)}$$

$$H_a : \text{var}(u_{it}) \neq 0 \Leftrightarrow H_0 : \text{regresion de panel}$$

El estadístico de contrastes es el LM con una distribución de una χ^2 .

Si el valor del test es bajo (p-valor mayor de 0.05) la hipótesis nula se confirma y es mejor MCO. Si el valor del test es alto (p-valor menor de 0.05) la hipótesis nula se rechaza y es mejor elegir un modelo de datos de panel.

- A la tercera pregunta se la puede contestar utilizando el contraste de Hausman, cuya hipótesis nula y alternativas son las siguientes:

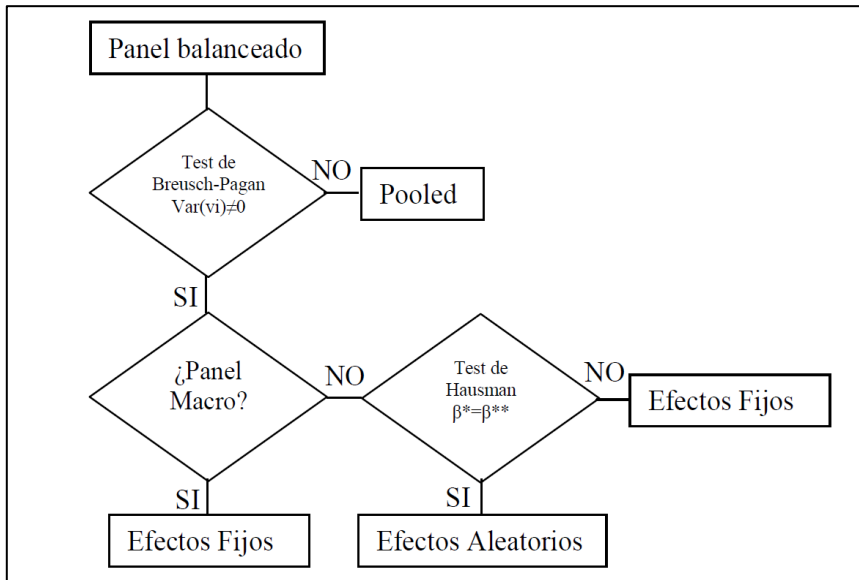
$H_0 : \text{corr}[z_i, x_{it}] = 0$ \Rightarrow estimadores MCO del modelo de efectos fijos y los mcg son consistentes, pero los primeros son ineficientes. \Rightarrow **coeficientes MEA = MEF**
 \Rightarrow **utilice efectos aleatorios**

$H_1 : \text{corr}[z_i, x_{it}] \neq 0$ \Rightarrow estimadores MCO del modelo de efectos fijos son consistentes, pero los MCG no. \Rightarrow **coeficientes MEA \neq MEF** \Rightarrow **utilice efectos fijos**

El mismo compara las estimaciones del modelo de efectos fijos y el de efectos aleatorios. Si encuentra diferencias sistemáticas se rechaza la hipótesis nula de igualdad, es decir se obtiene un valor del contraste alto y un p-valor bajo (menor de 0.05) y siempre que estemos medianamente seguros de la especificación, podremos

entender que continúa existiendo correlación entre el error y los regresores ($Cov(X_{it}, u_{it}) \neq 0$) y es preferible elegir el modelo de efectos fijos.

Gráfico 18. Esquema de especificación y estimación de modelos con datos de panel



5.- Análisis econométrico: resultados

Para analizar mediante datos de panel el efecto de las variables macroeconómicas seleccionadas en la brecha de salario entre hombre y mujeres en las 17 comunidades autónomas, se utilizará el programa estadístico gretl, con el cual es posible modelizar usando la técnica de datos de panel. Se trata de estimar el siguiente modelo;

$$D_lsal_{it} = f(\text{conjunto de variables explicativas}) + u_{it} \quad (10)$$

Donde D_lsal es la brecha salarial calculada como la diferencia salarial entre hombres y mujeres (el salario de hombres y mujeres se ha transformado tomando logaritmos neperianos). Dicha variable es la variable dependiente a explicar.

Las variables explicativas consideradas para explicar la brecha salarial entre hombre y mujeres son las siguientes

Dparo _{it}	Brecha de paro entre hombre y mujeres
P_Industria _{it}	Industria como porcentaje del PIB
P_Servicio _{it}	Servicio como porcentaje del PIB
A_ med educ _{it}	Años medio educación

El modelo finalmente especificado es el siguiente:

$$D\ln(\text{sal}_{it}) = \beta_0 + \beta_1 \text{Dparo}_{it} + \beta_2 \text{P_Industria}_{it} + \beta_3 \text{P_servicio}_{it} + \beta_4 \text{A_med_edu}_{it} + u_{it}$$

$i = 1, 2, \dots, 17$ (11)

$t = 2008, \dots, 2014$

Siguiendo el esquema planteado anteriormente (Gráfico 18), partimos de la estimación de la Regresión agrupada (pooled MCO) y a continuación, realizamos el contraste de Breusch-Pagan.

Tabla 1. Estimación MCO del modelo: Regresión Agrupada

Modelo 1_1:				
MCO combinados, utilizando 119 observaciones				
Se han incluido 17 unidades de sección cruzada				
Largura de la serie temporal = 7				
Variable dependiente: <u>Dl_SAL</u>				
	Coefficiente	Desv. Típica	Estadístico t	valor p
const	-0.389384	0.121921	-3.194	0.0018 ***
Dparo	0.00718651	0.00181885	3.951	0.0001 ***
p_Industria	0.00718612	0.00233720	3.075	0.0026 ***
P_Servicios	0.00292967	0.00178764	1.639	0.1040
A_med_educ	0.0204566	0.0115265	1.775	0.0786 *
Media de la vble. dep.	0.167116	D.T. de la vble. dep.	0.053086	
Suma de cuad. residuos	0.133936	D.T. de la regresión	0.034276	
R-cuadrado	0.597224	R-cuadrado corregido	0.583092	
F(4, 114)	42.25901	Valor p (de F)	1.08e-21	
Log-verosimilitud	235.1225	Criterio de Akaike	-460.2450	
Criterio de Schwarz	-446.3494	Crit. de Hannan-Quinn	-454.6025	
rho	0.625754	Durbin-Watson	0.609509	

Contraste de Breusch-Pagan:

Estadístico de contraste de Breusch-Pagan:

LM = 95.4382 con valor p = prob(chi-cuadrado(1) > 95.4382) = 1.52577e-022
(Un valor p bajo es una indicación en contra de la hipótesis nula de que el modelo de MCO combinados es el adecuado, en favor de la alternativa de efectos aleatorios.)

Estadístico de contraste de Hausman:

H = 4.84157 con valor p = prob(Chi-cuadrado(4) > 4.84157) = 0.303944

Se tratar de contrastar La hipótesis nula es $Var(ui)=0$. Como se puede observar obtenemos un valor del estadístico LM de 95,43 y un p-valor bastante bajo, indicando el rechazo de la hipótesis nula en favor de la alternativa de un modelo de datos de panel sea de efecto fijo o de efecto aleatorio.

Teniendo en cuenta el resultado del contraste de Breusch-Pagan, la siguiente pregunta es ¿las estimaciones consistentes (efectos fijos) y las eficientes (efectos aleatorios) son significativamente distintas? Para ello aplicamos el contraste de Hausman, Resulta muy importante conocer si el modelo adecuado para analizar nuestros datos es el de efectos fijos o el de efectos aleatorios Bajo la hipótesis nula de que se cumplen los supuestos del modelo de Efectos Aleatorios, ambos estimadores, el de efectos fijos y el de efectos aleatorios, deben ser similares ambos son consistentes. El contraste compara los coeficientes estimables de los regresores que varían con el tiempo. El estadístico de contraste mide la “distancia” entre ambas estimaciones: si es “grande” se rechaza H_0 .

Tabla 2.

```

Modelo 1_2:
Efectos aleatorios (MCG), utilizando 119 observaciones
Se han incluido 17 unidades de sección cruzada
Largura de la serie temporal = 7
Variable dependiente: Dl_SAL

-----
                Coeficiente   Desv. Típica      z      valor p
-----
const          -0.295306       0.131612        -2.244   0.0248  **
Dparo          0.00399801       0.00154140      2.594   0.0095  ***
p_Industria    0.0110199        0.00320874      3.434   0.0006  ***
P_Servicios    0.00510627       0.00224614      2.273   0.0230  **
A_med_educ     -0.00395084       0.0184650       -0.2140  0.8306

Media de la vble. dep.  0.167116   D.T. de la vble. dep.  0.053086
Suma de cuad. residuos  0.143353   D.T. de la regresión   0.035307
Log-verosimilitud      231.0796   Criterio de Akaike     -452.1591
Criterio de Schwarz    -438.2635   Crit. de Hannan-Quinn  -446.5165

Varianza 'entre' (between) = 0.000818138
Varianza 'dentro' (Within) = 0.000512384
theta usado para quasi-demeaning (cuasi-centrado de los datos) = 0.713432
corr(y,yhat)^2 = 0.568992

Contraste conjunto de los regresores (excepto la constante) -
  Estadístico de contraste asintótico: Chi-cuadrado(4) = 54.566
  con valor p = 4.0056e-011

Contraste de Breusch-Pagan -
  Hipótesis nula: Varianza del error específico a la unidad = 0
  Estadístico de contraste asintótico: Chi-cuadrado(1) = 95.4382
  con valor p = 1.52577e-022

Contraste de Hausman -
  Hipótesis nula: Los estimadores de MCG son consistentes
  Estadístico de contraste asintótico: Chi-cuadrado(4) = 4.84157
  con valor p = 0.303944

```

Como se puede observar los resultados obtenidos en el contraste de Hausman en donde se prueba la hipótesis nula de que los estimadores son consistentes se obtiene un p-valor de 0,3039, por lo que no se rechaza la hipótesis nula y se concluye que mediante efectos aleatorios los estimadores son consistentes por lo que efectos aleatorios sería la mejor opción.

El problema de estimar usando efectos fijos o aleatorias es que puede existir heteroscedasticidad o auto correlación entre las variables explicativas de manera que la inferencia puede estar mal hecha. Gretl no posee un contraste en datos de panel que permita analizar la existencia o no de heteroscedasticidad o auto correlación, lo que derivaría en un problema al momento de definir a las variables independientes como significativas o no.

Para solucionar este problema es necesario usar la estimación por desviaciones típicas robustas, de manera que la inferencia sea correcta y se pueda confiar en la significancia o no de una variable. Es en base a esta última estimación que se procederá a explicar las betas de cada variable y lo que cada una implicaría en el modelo. Los resultados de la estimación por desviaciones típicas robustas se muestran en la ilustración

Una vez que hemos inclinado por los estimadores consistentes los efectos aleatorios, a continuación hemos probado varias especificaciones que se muestran a continuación:

```

Modelo 1_3:
Efectos aleatorios (MCG), utilizando 119 observaciones
Se han incluido 17 unidades de sección cruzada
Largura de la serie temporal = 7
Variable dependiente: Dl_SAL
Desviaciones típicas robustas (HAC)

```

	Coefficiente	Desv. Típica	z	valor p	
const	-0.270854	0.134041	-2.021	0.0433	**
Dparo	0.00553902	0.000859208	6.447	1.14e-010	***
p_Industria	0.00422954	0.00113252	3.735	0.0002	***
A_med_educ	0.0297646	0.0111917	2.660	0.0078	***

```

Media de la vble. dep. 0.167116 D.T. de la vble. dep. 0.053086
Suma de cuad. residuos 0.149800 D.T. de la regresión 0.035936
Log-verosimilitud 228.4623 Criterio de Akaike -448.9245
Criterio de Schwarz -437.8080 Crit. de Hannan-Quinn -444.4105

Varianza 'entre' (between) = 0.000749206
Varianza 'dentro' (Within) = 0.00051749
theta usado para quasi-demeaning (cuasi-centrado de los datos) = 0.700314
corr(y,yhat)^2 = 0.549607

Contraste conjunto de los regresores (excepto la constante) -
Estadístico de contraste asintótico: Chi-cuadrado(3) = 60.9618
con valor p = 3.66212e-013

Contraste de Breusch-Pagan -
Hipótesis nula: Varianza del error específico a la unidad = 0
Estadístico de contraste asintótico: Chi-cuadrado(1) = 87.3548
con valor p = 9.06985e-021

Contraste de Hausman -
Hipótesis nula: Los estimadores de MCG son consistentes
Estadístico de contraste asintótico: Chi-cuadrado(3) = 9.32604
con valor p = 0.0252558

```

La interpretación de los coeficientes estimados es la siguiente:

Dparo: Ante el incremento de una unidad porcentual en la brecha de paro entre hombres y mujeres, la brecha salarial incrementaría en un 0,55%.

P_Industria: Ante un incremento de una unidad porcentual del peso que tiene la industria sobre el PIB, la diferencia salarial incrementaría un 0,44%.

A_med_educ: Ante un incremento de un año medio de educación, la brecha salarial incrementaría en 2,97%.

Podemos decir la educación influye en la brecha salarial, que la diferencia en paro afecta a la brecha salarial, también lo hace el peso de la industria en el PIB. Pero ¿Cuál es el efecto sobre la brecha salarial si consideramos la segregación laboral en los sectores de industria y servicios?

Veamos que ocurre cuando incluimos en sector servicios y seguidamente el industrial:

```

Modelo 1_4:
Efectos aleatorios (MCG), utilizando 119 observaciones
Se han incluido 17 unidades de sección cruzada
Largura de la serie temporal = 7
Variable dependiente: Dl_SAL
Desviaciones típicas robustas (HAC)

```

	Coefficiente	Desv. Típica	z	valor p
const	-0.310214	0.153791	-2.017	0.0437 **
Dparo	0.00593334	0.00110668	5.361	8.26e-08 ***
P_Servicios	-0.00201008	0.000734148	-2.738	0.0062 ***
A_med_educ	0.0491467	0.0103823	4.734	2.20e-06 ***

```

Media de la vble. dep. 0.167116 D.T. de la vble. dep. 0.053086
Suma de cuad. residuos 0.159638 D.T. de la regresión 0.037097
Log-verosimilitud 224.6776 Criterio de Akaike -441.3551
Criterio de Schwarz -430.2386 Crit. de Hannan-Quinn -436.8410

Varianza 'entre' (between) = 0.000753716
Varianza 'dentro' (Within) = 0.000563011
theta usado para quasi-demeaning (cuasi-centrado de los datos) = 0.689481
corr(y,yhat)^2 = 0.528507

Contraste conjunto de los regresores (excepto la constante) -
Estadístico de contraste asintótico: Chi-cuadrado(3) = 56.0448
con valor p = 4.10938e-012

Contraste de Breusch-Pagan -
Hipótesis nula: Varianza del error específico a la unidad = 0
Estadístico de contraste asintótico: Chi-cuadrado(1) = 82.9633
con valor p = 8.35905e-020

Contraste de Hausman -
Hipótesis nula: Los estimadores de MCG son consistentes
Estadístico de contraste asintótico: Chi-cuadrado(3) = 13.3745
con valor p = 0.00389298

```

Primero vemos que el p-valor del contraste de Hausman es inferior al 0,05 por lo que se rechaza la hipótesis nula y se concluye que mediante efectos fijos los estimadores son insesgados por lo que efectos fijos sería la mejor opción.

Vemos que el coeficiente de servicios tiene un valor negativo, es decir que cuando el peso de los servicios sobre el PIB aumenta, la brecha de salarios entre hombres y mujeres disminuye. Esto podría explicarse, en parte, por la segregación por sexos y sectores en el mercado laboral. En el sector servicios es donde se concentran la gran mayoría de las mujeres en el mercado laboral, por tanto, si partimos de una situación en la que los salarios de los hombres son en media mayores que los de las mujeres y el peso del sector servicios aumenta, cabe pensar que las mujeres se beneficiaran más que los hombres de ese incremento en el peso y por tanto la brecha debería reducirse.

Modelo 2_1:
Efectos fijos, utilizando 119 observaciones
 Se han incluido 17 unidades de sección cruzada
 Largura de la serie temporal = 7
 Variable dependiente: D1_SAL
 Desviaciones típicas robustas (HAC)

	Coefficiente	Desv. Típica	Estadístico t	valor p
const	-0.356060	0.208525	-1.708	0.1071
Dparo	0.00499604	0.000709932	7.037	2.81e-06 ***
p_Industria	0.0109146	0.00475857	2.294	0.0357 **
A_med_educ	0.0269934	0.0147517	1.830	0.0860 *

Media de la vble. dep. 0.167116 D.T. de la vble. dep. 0.053086
 Suma de cuad. residuos 0.051231 D.T. de la regresión 0.022748
 R-cuadrado MCVF (LSDV) 0.845936 R-cuadrado 'intra' 0.208648
 Log-verosimilitud 292.3025 Criterio de Akaike -544.6050
 Criterio de Schwarz -489.0225 Crit. de Hannan-Quinn -522.0347
 rho 0.120394 Durbin-Watson 1.489852

Contraste conjunto de los regresores (excepto la constante) -
 Estadístico de contraste: $F(3, 16) = 18.8461$
 con valor p = $P(F(3, 16) > 18.8461) = 1.67704e-005$

Contraste robusto de diferentes interceptos por grupos -
 Hipótesis nula: Los grupos tienen un intercepto común
 Estadístico de contraste: Welch $F(16, 38.1) = 10.54$
 con valor p = $P(F(16, 38.1) > 10.54) = 1.75747e-009$

Si analizamos de la misma forma que se ha hecho con anterioridad, se nos presenta la situación opuesta ya que el coeficiente del peso de la industria en el PIB es positivo. Vemos que el incremento del peso de la industria sobre el PIB debería incrementar la brecha salarial. Esto también concuerda con la teoría de la segregación al igual que en el caso del sector servicios. El sector industrial está más poblado de hombres que de mujeres. Por ello, si el peso de este sector incrementa, entonces los hombres se beneficiarían en media más que las mujeres y por tanto la brecha salarial entre hombres y mujeres aumentaría.

El efecto fijo en el modelo puede explicar el 84,56% de la diferencia de salarios.

```

Modelo 1_5:
Efectos aleatorios (MCG), utilizando 119 observaciones
Se han incluido 17 unidades de sección cruzada
Largura de la serie temporal = 7
Variable dependiente: D1_SAL
Desviaciones típicas robustas (HAC)

```

	Coefficiente	Desv. Típica	z	valor p	
const	-0.492065	0.198661	-2.477	0.0133	**
Dparo	0.00361518	0.00168722	2.143	0.0321	**
p_Industria	0.0114151	0.00404219	2.824	0.0047	***
P_Servicios	0.00611252	0.00326232	1.874	0.0610	*
A_med_educ	0.00724427	0.0160951	0.4501	0.6526	
dt_2	-0.0136669	0.00647097	-2.112	0.0347	**
dt_3	-0.0230666	0.00925843	-2.491	0.0127	**
dt_4	-0.0185037	0.0116146	-1.593	0.1111	
dt_5	0.000469717	0.0129653	0.03623	0.9711	
dt_6	-0.00634730	0.0124132	-0.5113	0.6091	
dt_7	-0.0387061	0.0106213	-3.644	0.0003	***

```

Media de la vble. dep. 0.167116 D.T. de la vble. dep. 0.053086
Suma de cuad. residuos 0.119230 D.T. de la regresión 0.033073
Log-verosimilitud 242.0431 Criterio de Akaike -462.0862
Criterio de Schwarz -431.5158 Crit. de Hannan-Quinn -449.6725

Varianza 'entre' (between) = 0.000840531
Varianza 'dentro' (Within) = 0.00035563
theta usado para quasi-demeaning (cuasi-centrado de los datos) = 0.761257
corr(y,yhat)^2 = 0.64145

Contraste conjunto de los regresores (excepto la constante) -
Estadístico de contraste asintótico: Chi-cuadrado(4) = 36.6927
con valor p = 2.08392e-007

```

```

Contraste de Hausman -
Hipótesis nula: Los estimadores de MCG son consistentes
Estadístico de contraste asintótico: Chi-cuadrado(4) = 5.15188
con valor p = 0.272065

```

Se rechaza el contraste de Hausman por tanto se utilizan efectos aleatorios.

En el “Modelo 1_5” se muestra un análisis de sensibilidad en que hemos incluido las variables temporales para ver los efectos que puedan darse año a año, es decir, efectos que provengan de la propia evolución de la economía, no alteran el signo de los coeficientes. Esta estimación es consistente con el grafico 10 ya que hay más diferencia los primeros años que al final del periodo tal y como indica la significatividad de los parámetros dt_2, dt_3 y dt_7. Insistimos que aquí no analizamos discriminación sino simplemente usamos unas variables económicas para analizar la evolución de la diferencia de salarios.

6.- Conclusiones

Ha habido un efecto rebote en las tasas de paro ya que se han reducido con el transcurso de la recesión y han vuelto a crecer en el periodo final de esta. Lo mismo ha ocurrido con los

salarios, pero de forma inversa, es decir, primero has aumentado y al final tendieron a decrecer. Ello ha ocurrido de forma diferente entre las regiones.

La brecha salarial ha aumentado durante la mayor parte de la crisis a razón de que el incremento de paro se ha debido a la pérdida de los empleos del sector de la construcción en el cual predominan los hombres, expulsó a los asalariados masculinos que cobraban un menor salario lo que en términos relativos causó un incremento de la brecha salarial. En los últimos dos años ésta ha disminuido.

Hemos realizado un análisis econométrico con el que explicamos hasta el 85% de la brecha con datos de panel en base al cual concluimos que en la brecha salarial influye el peso del sector industrial y el sector servicios sobre el PIB, los años de educación y las tasas de paro. Este último es influyente porque los trabajadores que sufren más paro cobran menos y generalmente los trabajos que requieren menos cualificación tienen más paro ya que hay más competencia. Los coeficientes de los datos afectan de la forma esperada, es decir, un incremento en el peso de la industria, o en la brecha de paro, incrementa la brecha salarial y un incremento en el sector servicios lo disminuye por lo que la teoría concuerda con los datos. Ello indica que la estructura productiva es importante.

Hay un 15% de la brecha salarial que nuestro modelo no explica. Ello puede ser debido a que omitimos variables relevantes (ya sea porque no existen para España y CCAA o no sean cuantificables) o podría ser discriminación.

7.- Bibliografía

Cantó, Olga, Inmaculada Cebrián, and Gloria Moreno. "Crisis y brecha de riesgo de pobreza por género." *Estudios de Economía Aplicada* 34.1 (2016).

De la Rica S, Rebollo-Sanz Y (2017) Gender differentials in unemployment ins and outs during the great recession in Spain. *De Economist* 165(1):67–69

DEL RÍO, Coral; VILLAR, Olga Alonso. Mujeres ante el empleo (y el desempleo) en el mercado laboral español, 1996-2013. *Areas. Revista Internacional de Ciencias Sociales*, 2014, no 33, p. 87-103.

García, A. N. (2017). Gender Differences in Unemployment Dynamics and Initial Wages over the Business Cycle. *Journal of Labor Research*, 38(2), 228-260

Guner N, Kaya E, Sánchez-Marcos V (2014) Gender gaps in Spain: policies and outcomes over the last three decades. *SERIEs* 5(1):61–103

Hill Carter, R., E. Griffiths William, and C. Lim Guay. "Principles of econometrics 4th ed. John Wiley & Sons Inc." (2012)

Huertas, I. P. M., Ramos, R., & Simon, H. (2017). Regional differences in the gender wage gap in Spain. *Social Indicators Research*, 134(3), 981-1008

Murillo, I. P., & Simón, H. (2014). La Gran Recesión y el diferencial salarial por género en España. *Hacienda Pública Española*, 208(1), 39-76.

Peña-Boquete Y (2014) Have the economic crises reduced the gender gap on the Spanish labour market? *Revue de l'OFCE* 2/2014 (N° 133):277-302