

Trabajo Fin de Máster  
Máster Universitario en Ingeniería Industrial

UN ANÁLISIS REGIONAL DE LA EFICIENCIA  
DEL EMPAREJAMIENTO LABORAL EN  
ESPAÑA

Autor: Francisco Javier Gordillo Montesino

Tutor: Fernando Núñez Hernández

Dpto. Organización Industrial y Gestión de Empresas I  
Escuela Técnica Superior de Ingeniería  
Universidad de Sevilla

Sevilla, 2019





Trabajo Fin de Máster  
Máster Universitario en Ingeniería Industrial

**UN ANÁLISIS REGIONAL DE LA EFICIENCIA DEL  
EMPAREJAMIENTO LABORAL EN ESPAÑA**

Autor:  
Fco.Javier Gordillo Montesino

Tutor:  
Fernando Núñez Hernández

Departamento de Organización Industrial y Gestión de Empresas I

Escuela Técnica Superior de Ingeniería

Universidad de Sevilla

Sevilla, 2019



Trabajo Fin de Máster: Un análisis regional de la eficiencia del emparejamiento laboral en España.

Autor: Fco. Javier Gordillo Montesino

Tutor: Fernando Núñez Hernández

El tribunal nombrado para juzgar el Proyecto arriba indicado, compuesto por los siguientes miembros:

Presidente:

Vocales:

Secretario:

Acuerdan otorgarle la calificación de:

Sevilla, 2019

El Secretario del Tribunal





# ÍNDICE

1. INTRODUCCIÓN.....	12
2. FUNDAMENTO TEÓRICO DEL EMPAREJAMIENTO LABORAL.....	15
2.1. Concepto de la función de emparejamiento.....	15
2.2. Especificaciones y propiedades matemáticas.....	16
2.3. Microfundamentos de la función agregada de emparejamiento.....	17
2.3.1. El microfundamento basado en el desajuste laboral o <i>mismatch</i> .....	18
2.3.2. El microfundamento de los fallos de coordinación.....	18
2.3.3. La intensidad de búsqueda y los salarios de reserva como microfundamentos del emparejamiento.....	19
2.3.4. El microfundamento basado en la clasificación.....	20
2.3.5. El microfundamento stock-flujo.....	21
2.3.6. Agregación sobre mercados diferentes.....	23
2.4. Problemas de agregación.....	23
2.4.1. Problemas de agregación temporal.....	23
2.4.2. Problemas de agregación espacial.....	24
3. METODOLOGÍA ECONOMETRICA: DATOS DE PANEL.....	26
3.1. Introducción a la metodología Datos de Panel.....	26
3.2. Modelos para la estimación de Datos de Panel.....	27
3.2.1. Modelo de datos de panel agrupados.....	28
3.2.2. Modelo de efectos fijos o estimación intra-grupos ( <i>within estimation</i> ).....	28
3.2.3. Modelo entre-grupos ( <i>between estimation</i> ).....	30
3.2.4. Modelo de efectos aleatorios.....	30
4. DESCRIPCIÓN DE LOS DATOS EMPLEADOS SOBRE EL EMPAREJAMIENTO LABORAL EN ESPAÑA A NIVEL REGIONAL.....	32
4.1. Análisis descriptivo de las variables laborales. Series trimestrales. Período 2006-2018. 32	
4.1.1. La <i>Estadística de Empleo</i> del Servicio Público de Empleo Estatal.....	33
4.2. Evolución anual del mercado de trabajo en España.....	33
4.3. Evolución trimestral del stock de demandantes por CCAA.....	36
4.4. Evolución trimestral del stock de vacantes por CCAA.....	38
4.5. Evolución trimestral del flujo de nuevos demandantes por CCAA.....	41
4.6. Evolución trimestral del flujo de nuevas vacantes por CCAA.....	43
4.6.1. Re-escalamiento del flujo de nuevas vacantes por CCAA.....	45
4.7. Evolución trimestral del flujo de colocaciones por CCAA.....	46



4.8.	Descripción de las variables laborales en España y por CCAA. Periodo 2006:T1–2018:T4. ....	48
4.8.1.	Descripción de las variables laborales por CCAA. 2006:01 – 2018:04. ....	51
4.9.	Descripción de otras variables relevantes del mercado del trabajo en España. 2006:01 – 2018:04. ....	55
5.	ESTIMACIÓN DEL MODELO ECONOMETRICO Y PRINCIPALES RESULTADOS. ....	58
5.1.	Comparativa entre el estimador de efectos fijos y el de efectos aleatorios (Test de Hausman). ....	59
5.2.	Estimación agrupada de la función de emparejamiento a partir de la metodología de datos de panel. ....	60
5.3.	Estimación de la función de emparejamiento aplicando el estimador de efectos fijos. ....	61
6.	CONCLUSIONES. ....	70
7.	REFERENCIA BIBLIOGRÁFICA. ....	73

## ÍNDICE DE FIGURAS

Figura 4.1: Evolución trimestral del mercado de trabajo en España. Periodo 2006:T1-2018:T4, 34.

Figura 4.2: Curva de Beveridge del mercado de trabajo en España. Periodo 2006:T1-2018:T4, 35.

Figura 4.3: Curva de Beveridge del mercado de trabajo en Andalucía. Periodo 2006:T1-2018:T4, 36.

Figura 4.4: Evolución trimestral del stock de demandantes por CCAA. Periodo 2006:T1-2018:T4, 38.

Figura 4.5: Evolución trimestral del stock de vacantes por CCAA. Periodo 2006:T1-2018:T4, 41.

Figura 4.6: Evolución trimestral del flujo de nuevos demandantes por CCAA. Periodo 2006:T1-2018:T4, 43.

Figura 4.7: Evolución trimestral del flujo de nuevas vacantes por CCAA. Periodo 2006:T1-2018:T4, 44.

Figura 4.8: Evolución trimestral del flujo de colocaciones por CCAA. Periodo 2006:T1-2018:T4, 47.

Figura 5.1: Estimación de la función de emparejamiento a partir del modelo general de datos de panel. Periodo 2009:T1-2018:T4, 60.

Figura 5.2. Estimación de datos de panel aplicando el estimador de efectos fijos. Periodo 2009:T1-2018:T4, 67.

Figura 5.3. Residuos de las series reales y las series estimadas frente al predictor. Periodo 2009:T1-2018:T4, 68.

Figura 5.4. Efectos fijos frente al flujo de colocaciones o predictor en escala logarítmica (base 10). Periodo 2009:T1-2018:T4, 69.

## ÍNDICE DE TABLAS

Tabla 4.1: Descripción de las variables laborales en España. Periodo 2006:T1-2018:T4, 50.

Tabla 4.2: Descripción del stock de demandantes por CCAA. Periodo 2006:T1-2018:T4, 51.

Tabla 4.3: Descripción de contratos iniciales por CCAA. Periodo 2006:T1-2018:T4, 52.

Tabla 4.4: Descripción de la tasa de emparejamiento de puestos por CCAA. Periodo 2006:T1-2018:T4, 53.

Tabla 4.5: Descripción de la tasa de emparejamiento de trabajadores por CCAA. Periodo 2006:T1-2018:T4, 54.

Tabla 4.6: Descripción del grado de tensión por CCAA. Periodo 2006:T1-2018:T4, 55.

Tabla 4.7: Estructura de la generación de empleo. Valores medios trimestrales. Periodo 2006:T1-2018:T4, 56.

Tabla 4.8: Estructura del mercado de trabajo y temporalidad en el stock de ocupados. Periodo 2006:T1-2018:T4, 57.

Tabla 5.1: Resultados de la función de emparejamiento estimada por el modelo de efectos fijos. Periodo 2009:T1-2018:T4, 62.

## **1. INTRODUCCIÓN.**

En el mercado laboral español, durante el período 2008-2018, se experimentan principalmente dos situaciones. Entre 2008 y 2013 experimentamos una crisis económica, caracterizada por el registro de una de las mayores tasas de desempleo conocidas por nuestro país, alrededor del 25 % de la población activa al inicio de 2013. Este último año se considera un punto de inflexión, ya que es a partir de 2014 cuando comienza un periodo de expansión económica que dura hasta 2018 –el año 2019 muestra algunos síntomas de desaceleración económica. Efectivamente, en el periodo 2014-2018 se produce una fuerte generación de empleo que reduce los niveles de paro en 10 puntos porcentuales respecto a 2013 (la tasa de paro es del 14,4% en el último trimestre de 2018).

Según las teorías actuales sobre el funcionamiento del mercado de trabajo, la existencia de desempleo se debe a desequilibrios entre la oferta y la demanda de trabajo y a la existencia de fricciones y de heterogeneidad en los mercados de trabajo; algunas de ellas se discuten en el modelo o teoría de búsqueda y emparejamiento laboral desarrollada por Petrongolo y Pissarides (2001). La movilidad laboral imperfecta de los desempleados, la heterogeneidad de los trabajadores y la heterogeneidad de los puestos, las fricciones procedentes de las políticas gubernamentales (por ejemplo, en materia de prestaciones por desempleo), o la existencia de información incompleta de las vacantes disponibles, son fricciones que afectan negativamente a la generación de empleo. Según el modelo del desempleo de Pissarides, base teórica de este Trabajo Fin de Máster, la tasa de desempleo en una región puede alcanzar un valor significativo si existe un lento y costoso proceso de emparejamiento entre la oferta y la demanda de trabajo. Además, dicha ineficiencia en la generación de empleo no tiene por qué afectar a todas las regiones del país por igual, pudiendo haber algunas regiones más eficientes que otras.

El presente trabajo pretende analizar el proceso de emparejamiento en el mercado de trabajo español a partir de las variables laborales que lo explican; a saber, el flujo de colocaciones (nuevos contratos), el flujo y el stock de desempleados, y el flujo y el stock de puestos vacantes. Estas variables se analizan adoptando una visión desagregada, territorial y sectorialmente. En primera instancia, se describen estas variables laborales a nivel nacional. Posteriormente, se muestra el proceso evolutivo de cada una de ellas en cada una de las 17 CCAA que componen el territorio nacional. Finalmente, se analizan ciertas características adicionales del emparejamiento laboral español, como por ejemplo el porcentaje de contratos

generados en los diferentes sectores de actividades económicas (agricultura, industrial, construcción y servicios).

El proceso de emparejamiento laboral (y su eficiencia) se puede representar mediante una expresión matemática propuesta por Pissarides (2000), conocida con el nombre de Función Agregada de Emparejamiento; función que, en su versión más simple, permite determinar el número de colocaciones o contratos iniciales generados en una región concreta, y en un intervalo de tiempo determinado, a partir del stock de vacantes y de desempleados existente en dicha región al inicio del intervalo de tiempo considerado. Autores como Coles y Smith (1998) y Petrongolo y Pissarides (2001) proponen expresiones más complejas y realistas de la función de emparejamiento, incluyendo variables adicionales en la función y proponiendo determinadas especificaciones matemáticas. Así, estos autores expresan el flujo de colocaciones en un intervalo de tiempo como dependiente de cuatro variables laborales: por un lado, el stock de demandantes ( $U$ ) y el stock de vacantes ( $V$ ), ambos stocks referidos al comienzo del periodo analizado, y por otro lado el flujo de nuevos buscadores de empleo ( $uf$ ) y el flujo de nuevas vacantes ( $vf$ ) que surgen a lo largo del intervalo de tiempo considerado.

Además de los artículos citados anteriormente, entre las referencias bibliográficas en las que se fundamenta este Trabajo Fin de Máster se encuentran también las aportaciones de Antolín (1994), de Coles y Petrongolo (2003) y de Nuñez y Usabiaga (2007). Estos últimos autores realizan una revisión de los fundamentos teóricos y de la literatura empírica existente sobre la función de emparejamiento, con especial énfasis en el caso español.

Los datos empleados en este trabajo provienen de las series estadísticas publicadas por diversos organismos públicos. Concretamente, para la obtención del stock de demandantes ( $U$ ), del flujo de nuevos desempleados ( $uf$ ), del flujo de nuevas vacantes generadas ( $vf$ ) y del stock de contratos iniciales se ha recurrido a la *Estadística de Empleo* del Servicio Público de Empleo Estatal (SEPE). Al mismo tiempo, por la falta de representatividad de algunas de las variables laborales publicadas en el organismo anterior, en la sección 4 del presente trabajo, se proponen ciertas correcciones de los datos con el objetivo de obtener series temporales representativas que sean capaces de reflejar mejor la realidad del mercado laboral español. Por su parte, la serie de datos correspondiente al stock de vacantes ( $V$ ) se ha extraído, hasta el año 2013, de la *Encuesta de Coyuntura Laboral* (ECL), publicada por el Ministerio de Trabajo, Migraciones y Seguridad Social, y, a partir de 2013, de la *Encuesta Trimestral del Coste Laboral* (ETCL), publicada por el Instituto Nacional de Estadística (INE). Finalmente, hemos empleado otras series procedentes del INE, como, por ejemplo,

las referentes a la estructura de la población activa (ocupados y desempleados) y que, en concreto, se pueden encontrar en su *Encuesta de Población Activa* (EPA).

A la hora de analizar la eficiencia del emparejamiento laboral en el mercado de trabajo, se pueden emplear diversas metodologías econométricas (modelos de regresión lineal, de regresión no lineal, modelos logit/probit, de regresión con datos de panel o econométricos dinámicos, entre otros). La técnica en la que se fundamenta este Trabajo de Fin de Máster es Datos de Panel, la cual se caracteriza por estudiar el comportamiento de  $i$  individuos durante  $N$  periodos de tiempo equidistantes (cuatrimestres, años, etc); la particularidad de dicha metodología, a diferencia de otras, como, por ejemplo, del modelo de datos de corte transversal, reside en que la muestra analizada en los diferentes momentos de tiempo u observaciones  $t$  no varía. Esta propiedad descrita despierta especial interés en los economistas laborales, ya que proporciona una información completa de los individuos, en nuestro caso, las comunidades autónomas españolas; por un lado, captura la evolución de cada comunidad a lo largo de los diferentes periodos temporales y, por otro, ofrece una comparativa entre las regiones observadas en la misma unidad de tiempo. Al mismo tiempo, otra característica relevante de la metodología especificada es que permite controlar o modelizar la heterogeneidad inobservable de las diferentes unidades muestrales. Entre los modelos que se emplean para la estimación de un panel de datos destacan la regresión agrupada, el estimador de efectos fijos y el estimador de efectos aleatorios.

El presente trabajo se estructura de la siguiente forma: en primer lugar, en la sección 2, se expone el concepto de función de emparejamiento, las propiedades y especificaciones matemáticas de dicha función y los micro fundamentos en los que se basa. A continuación, en la sección 3 se introduce la metodología de datos de panel y los diferentes especificaciones o modelos que permiten estimarlo –entre ellos destaca el modelo de datos agrupados, el estimador de efectos fijos y el estimador de efectos aleatorios. La sección 4 ofrece un análisis descriptivo, a nivel nacional, de las series de las variables laborales influyentes en la función de emparejamiento, así como, la evolución de cada una de ellas por comunidad autónoma; destacar que, ciertas series se han ajustado/reescalado debido a su falta de representatividad estadística. En la sección 5 se analizan los resultados obtenidos de la estimación del modelo econométrico. Finalmente, la sección 6 describe las principales conclusiones deducidas de este Trabajo Fin de Máster.

## 2. FUNDAMENTO TEÓRICO DEL EMPAREJAMIENTO LABORAL.

### 2.1. Concepto de la función de emparejamiento.

En la línea de las aportaciones teóricas de Pissarides (2000) y Shimer y Smith (2001) –véase Núñez y Usabiaga (2007) para la economía española–, cuando los puestos ofrecidos muestran requerimientos diferentes (en términos de cualificación, localización geográfica, salario, edad, etc.), y los trabajadores poseen diferentes características, el emparejamiento laboral se puede convertir en un proceso lento y costoso, en el que cada trabajador debe esforzarse para encontrar el puesto que más se adapta a sus características, y a la inversa. Además, esta falta de homogeneidad entre las vacantes, por un lado, y los desempleados, por otro, es solo una de las fricciones que se pueden observar en los mercados de trabajo reales. Otros factores como las fluctuaciones del ciclo económico, las diferentes políticas gubernamentales de apoyo al empleo, el desconocimiento por parte del desempleado de las vacantes disponibles (información incompleta), la existencia de puestos con una rotación de personal considerable o la existencia del seguro por desempleo son otros de los muchos factores que parecen explicar la existencia de un desempleo permanente en la economía.

La necesidad de obtener una respuesta a la existencia de desequilibrio laboral y de desempleo permanente ha originado un conjunto de teorías no competitivas sobre el funcionamiento del mercado de trabajo, dejando de lado el enfoque clásico walrasiano, por el cual se establece el vaciado del mercado de trabajo (igualdad oferta-demanda) una vez alcanza su punto de equilibrio. El objeto del presente Trabajo de Fin de Máster es analizar la teoría no competitiva de búsqueda y emparejamiento (*search and matching models*) propuesta por el nobel de economía Pissarides –véase Pissarides (2000). Como veremos a continuación, el elemento central de esta teoría sobre el mercado de trabajo es la existencia de una función de emparejamiento a nivel agregado que admite varios microfundamentos. Dicha función pretende formalizar matemáticamente la tecnología que subyace en el proceso de emparejamiento laboral, proceso que pone en contacto a los desempleados (y al resto de buscadores de empleo) con las ofertas de trabajo publicadas, expresando así, el número de emparejamientos formados en cualquier intervalo de tiempo. La función de emparejamiento, en su forma más simple viene definida de la siguiente forma:

$$M = m(U, V) \tag{2.1}$$

donde  $M$  representa el número de emparejamientos en un diferencial de tiempo  $dt$  (si el modelo está en versión continua), el cual depende del stock de vacantes disponibles  $V$  y del stock de desempleados  $U$  en dicho instante. Si el modelo se discretiza, las colocaciones en un intervalo finito de tiempo ( $\Delta t$ ) dependerán de los desempleados y las vacantes que se encuentran dispuestos a emparejarse desde el comienzo de dicho intervalo temporal.

## 2.2. Especificaciones y propiedades matemáticas.

En el presente proyecto, la función de emparejamiento se basa en la función de producción de Cobb-Douglas –especificación que es habitual en la literatura empírica de este campo (Petrongolo y Pissarides, 2001)–, que adopta la siguiente forma:

$$M = AU^{\alpha_1}V^{\alpha_2} \quad 0 < \alpha_1, \alpha_2 < 1 \quad (2.2)$$

donde  $A$  expresa el factor tecnológico –conocido también con el nombre de productividad total de los factores (PTF)–, que representa factores no observados que favorecen el emparejamiento de los inputs  $U$  y  $V$ , como la información disponible, el salario de reserva, las nuevas tecnologías de la información existentes, etc.

Por otro lado,  $\alpha_1$  y  $\alpha_2$  son términos de elasticidad que hacen referencia a la variación del flujo de colocaciones al variar el stock de desempleados  $U$  o de vacantes  $V$ , respectivamente. Los rendimientos de escala constante se dan cuando se cumple:  $\alpha_1 + \alpha_2 = 1$  en la ecuación (2.2). Esta propiedad implica que, si demandantes y vacantes se incrementan en la misma proporción, el flujo de colocaciones aumentará en esa misma proporción. Si la suma de  $\alpha_1 + \alpha_2 > 1$ , se habla de rendimientos de escala crecientes, es decir, aumentando en la misma proporción los desempleados y las ofertas de trabajo, el flujo de colocaciones aumenta en una proporción mayor a la del stock de desempleados y vacantes. La situación opuesta sucede si  $\alpha_1 + \alpha_2 < 1$ , donde la variable de colocaciones aumenta en una relación menor a la de las variables de desempleados y vacantes.

Algunas evidencias de rendimientos de escala en la función de emparejamiento se observan en Petrongolo y Pissarides (2001). Autores como Pissarides (1986) y Layard et al. (1991) demuestran rendimientos de escala constantes en Reino Unido (UK). En el caso de España y Francia, los estudios realizados por Burda y Wyplosz (1994) muestran rendimientos de escala decrecientes en ambos casos. Un ejemplo de rendimientos de escala crecientes se aprecia en Alemania, a partir del análisis desarrollado por Münich et al. (1999).



Tomando como referencia la ecuación anterior, podemos establecer que un desempleado encuentra trabajo durante un período de tiempo de longitud la unidad con una probabilidad  $M/U$  –este ratio representa la tasa de emparejamiento de los trabajadores. De la misma forma, una vacante es cubierta con una probabilidad  $M/V$  –este ratio representa la tasa de emparejamiento de los puestos vacantes. La inversa de ambas tasas de emparejamiento, en un ambiente estacionario (sin perturbaciones), mide la duración media del desempleo y de las vacantes, respectivamente. Obviamente, las tasas anteriores pueden diferir entre trabajadores y entre puestos, dado que unos y otros presentan características individuales diferenciadas –existe heterogeneidad a ambos lados del mercado.

Otro indicador relevante es el grado de tensión existente en el mercado de trabajo ( $V/U$ ), el cual aumenta si existe un elevado número de puestos vacantes y escasos desempleados. Bajo estas circunstancias, cada empresa debe encontrar un candidato con las aptitudes idóneas entre la escasa oferta de trabajadores. En cambio, si el número de vacantes es pequeño y existen muchos demandantes, el grado de tensión disminuye, ya que para las empresas resulta más sencillo cubrir estas vacantes debido a la amplia oferta de mano de obra; en el caso de los trabajadores, por el contrario, se genera una situación de competencia y congestión.

### **2.3. Microfundamentos de la función agregada de emparejamiento.**

Aunque diversos modelos microeconómicos se pueden emplear para justificar la existencia de la función de emparejamiento agregada, y cada uno de ellos define una determinada tecnología del emparejamiento, no en todos ellos la función de emparejamiento se expresa de la forma empírica habitual de Cobb-Douglas. En cambio, si hay bastante consenso en la literatura a la hora de sugerir las variables que deberían de influir en la tasa de emparejamiento de trabajadores y de puestos. Estas variables se pueden clasificar en dos grupos. El primero de ellos engloba a todas aquellas tareas que los agentes (trabajadores o empresas) realizan durante la búsqueda de empleo, como elegir el número de compañías en las que aplicarán, seleccionar los canales de búsqueda, modificar sus métodos para mostrarse al mercado, etc. Por otro lado, el segundo grupo de variables explicativas engloba factores que no están relacionados, al menos directamente, con las decisiones individuales de búsqueda de empleo, siendo más bien variables del entorno del individuo.

### **2.3.1. El microfundamento basado en el desajuste laboral o *mismatch*.**

El término “*mismatch*” es un concepto empírico que mide el grado de desajuste que existe entre trabajadores y puestos vacantes. Los trabajadores muestran una serie de características (habilidades, formación, grupo de ocupación, sector de actividad de pertenencia, salario de reserva, lugar de residencia, etc.) que no siempre casan con los requerimientos de los puestos en términos de formación, remuneración, etc.

Las posibles diferencias existentes entre las habilidades que poseen los trabajadores y aquellas que requieren las compañías originan un aumento en el tiempo que se emplea para emparejar a un grupo de trabajadores con una serie de empresas dadas, ya que los agentes económicos tratan de encontrar un emparejamiento apropiado entre los posibles candidatos heterogéneos. A medida que evoluciona el mercado de trabajo, los diferentes sectores industriales y de actividad, y se producen avances tecnológicos, es necesario, por parte de los trabajadores formarse para poder adaptarse a las nuevas demandas emergentes de las empresas.

Existen también desajustes de localización entre los buscadores de empleo y las vacantes, también conocidos en la literatura como desequilibrios entre mercados locales, que influyen en el emparejamiento debido a la imperfecta movilidad laboral de los trabajadores. Para distinguir entre el desajuste de habilidades/formativo y de localización, se emplea un término adicional para hacer referencia a este último, denominado “*imbalance*” o desequilibrio laboral.

Si ambos términos explicados, tomasen un valor igual a cero, la función de emparejamiento no existiría ya que los trabajadores y las vacantes se emparejarían instantáneamente, considerándose el mercado laboral como un mercado perfecto.

### **2.3.2. El microfundamento de los fallos de coordinación.**

El primer modelo sobre la función de emparejamiento laboral debe su origen a la teoría de las bolas y las urnas. Las empresas desempeñan el rol de las urnas, mientras que los trabajadores representan las bolas. Una compañía se considera productiva si una bola se encuentra en su interior tras la ronda de emparejamiento correspondiente. La situación en la que el número de bolas y urnas sea el mismo, no garantiza un emparejamiento perfecto, debido a la ubicación aleatoria de las bolas en el interior de las urnas, de modo que se puede dar la situación de que determinadas urnas contengan más de una bola y otras queden vacías. A este caso se le denomina fallos de coordinación. En el contexto del mercado laboral, la existencia de fallos de coordinación implicaría que, durante un proceso de aplicación de

candidatos, algunas vacantes experimentarían saturación, existiendo otras en las cuales ningún trabajador aplicaría. La raíz de este problema de coordinación, que provoca un desequilibrio en el emparejamiento laboral y, por tanto, origina desempleo, es la falta de información sobre las acciones del resto de aspirantes.

La expresión más simplificada de este microfundamento parte del supuesto de que los desempleados (y el resto de los trabajadores en búsqueda de empleo), es decir, el stock  $U$ , conocen la localización de las vacantes existentes y aplican a esas ofertas. Si sucede que una oferta publicada recibe varias aplicaciones, se elige aleatoriamente a un aspirante, formando un emparejamiento; los no elegidos son devueltos a la “*pool*” de desempleados para así aplicar de nuevo. Por tanto, teniendo en cuenta las consideraciones descritas y, además, suponiendo un elevado stock de vacantes, la función de emparejamiento se expresa de la forma:

$$M = V(1 - e^{-U/V}) \quad (2.3)$$

La agregación de pequeñas fricciones a la ecuación (2.3), como pueden ser el desconocimiento de las vacantes disponibles por parte de los desempleados, las discrepancias entre las habilidades demandadas por las empresas y las individuales de cada trabajador, a lo que se le añade el desconocimiento de los desempleados de las vacantes que resultan acordes a sus aptitudes, o la inclusión de un factor de intensidad, permiten enriquecer la función de emparejamiento anterior.

### **2.3.3. La intensidad de búsqueda y los salarios de reserva como microfundamentos del emparejamiento.**

Existen diversas características de los individuos que influyen en la propia tecnología del emparejamiento y, por tanto, afectan a la función de emparejamiento agregada. La heterogeneidad del trabajador se puede introducir de forma explícita en la función de emparejamiento a través del factor intensidad de búsqueda; cada trabajador elige (variable de elección) qué intensidad de búsqueda quiere aplicar. Por tanto, se define la intensidad de búsqueda como el número de “unidades” de búsqueda suministradas por un individuo dado. Si un individuo  $i$  suministra  $s_i$  unidades de búsqueda y un individuo  $j$ ,  $s_j$  unidades, entonces en un intervalo pequeño de tiempo, el individuo  $i$  tiene  $s_i/s_j$  veces más probabilidad que el individuo  $j$  de encontrar una pareja. Las unidades de búsqueda son suministradas a un coste, por lo que cada trabajador tiene que elegir las de forma óptima para maximizar los

rendimientos netos de su búsqueda. Por tanto, diferentes individuos elegirán un número de unidades de búsqueda, dependiendo de sus costes, del coste de desempleo, y los rendimientos esperados del empleo. Aplicando el factor de intensidad de búsqueda en la ecuación (2.4), se obtiene una nueva expresión de la función de emparejamiento agregada

$$M = m(sU, V) \quad (2.4)$$

siendo  $s$  el promedio de unidades de búsqueda suministradas por un desempleado. De modo que, considerando la heterogeneidad en la intensidad de búsqueda, la tasa de emparejamiento para los trabajadores es  $s_i m(sU, V)/sU$ .

La distribución en las ofertas de salarios es otro medio a través del cual la heterogeneidad puede influir en los resultados de la función de emparejamiento. Según este enfoque, el individuo elige un salario de reserva o de referencia y rechaza aquellos salarios que están por debajo de dicho valor. En los modelos de equilibrio, ni la empresa ni el trabajador tendrán intención de formar un emparejamiento si el salario está por debajo del valor umbral definido. Naturalmente, si las características de los individuos difieren, los trabajadores elegirán diferentes salarios de reservas. Introduciendo el factor descrito, la función de emparejamiento resultante se observa en la ecuación (2.5).

$$M = [1 - G(R)]m(U, V) \quad (2.5)$$

Suponiendo un salario de reserva  $R_i$ , el trabajador aceptaría un salario  $\omega$  (asumiendo una distribución de probabilidad  $G(\omega)$  y conocida por los desempleados) ofertado por la empresa, siempre y cuando se cumpla la condición  $\omega \geq R_i$ . Dado que la distribución de probabilidad se considera no-lineal, lo que implicaría una función de emparejamiento más compleja, para simplificar se asume un salario de reserva promedio  $R$ .

#### **2.3.4. El microfundamento basado en la clasificación.**

Blanchard y Diamond (1994) parten de la suposición de que una oferta de trabajo publicada por una empresa recibe un gran número de aplicaciones o aspirantes al mismo tiempo, por lo que las compañías manifiestan ciertas preferencias sobre los aplicantes. La compañía clasifica a los aplicantes y le ofrece el trabajo a la persona que se encuentra en la primera posición del ranking, en base a las habilidades que ellos consideran apropiadas.

Supongamos que los desempleados están divididos en dos grupos, de corta duración y los de larga duración. Sea el número de desempleados de corta duración  $U^S$ , y el número de

desempleados de larga duración  $U^L$ ; supongamos además que ambos grupos compiten por los mismos puestos de trabajo, en estas condiciones, aquel colectivo que ha estado menos tiempo desempleado tiene mayores opciones de ocupar los puestos ofertados –los desempleados de larga duración presentan un efecto estigma que los perjudica. Por tanto, los desempleados de larga duración no provocan congestión a los de corta duración durante el proceso de búsqueda, y además solo consiguen aquellos trabajos no solicitados por los desempleados de corta duración. Si ahora, consideramos los fallos de coordinación entre los desempleados de corta duración y los de larga duración, ya que este último grupo desconoce las vacantes tomadas por los desempleados de corta duración, la función de emparejamiento resulta en la forma:

$$M = m(U^S + U^L, V) \quad (2.6)$$

Para esta situación, la tasa de emparejamiento para trabajadores de corta duración se define como  $m^S(U^S, V)/U^S$  mientras que, los desempleados de larga duración se emparejan con una probabilidad  $m(U^S + U^L, V)/U^L - m^S(U^S, V)/U^L$ .

Obsérvese que todos los modelos que venimos exponiendo implican, de una forma o de otra, que el emparejamiento laboral es un proceso que requiere de tiempo (y dinero), no es algo instantáneo.

### **2.3.5. El microfundamento stock-flujo.**

Durante los subapartados anteriores, se ha consolidado la suposición de que las personas en búsqueda de trabajo elegían una vacante de forma aleatoria y aplicaban a ella. En este fundamento de la función de emparejamiento, se descarta la hipótesis indicada y nos centraremos en el extremo opuesto, es decir, la no existencia de aleatoriedad en las aplicaciones de trabajo; el emparejamiento es fruto de un proceso dirigido, no aleatorio.

Coles (1994) y Coles y Smith (1998) consideran que los buscadores de trabajo tienen información completa sobre las vacantes disponibles y aplican simultáneamente a aquéllas en las que tienen una alta probabilidad de ser aceptados. Sea la constante  $\alpha$  la probabilidad de que un emparejamiento desempleado-vacante sea inaceptable. Una ronda de emparejamiento se produce en un punto de encuentro o lugar de mercado o “*marketplace*”. En dicho sitio de encuentro, todos los trabajadores observan a todas las vacantes y a la inversa, de modo que se acaban generando todos los emparejamientos productivos posibles, ninguna empresa y ningún trabajador que pudiera formar un emparejamiento productivo van a permanecer sin pareja. Una vez que la ronda de emparejamiento ha concluido, los

trabajadores y puestos sin emparejar tienen que esperar a que, en una siguiente ronda, entren en el sitio de mercado nuevas parejas potenciales para intentar emparejarse. Obsérvese que, a diferencia de como sucedía en la teoría de urnas y bolas, en este caso no existe falta de coordinación y aquellos trabajadores que permanecen desemparejados es porque no existen, entre las ofertadas en el mercado, vacantes adecuadas para ellos.

Las suposiciones de Coles y Smith contemplan una característica realista del mercado laboral: una persona en búsqueda de trabajo realiza un análisis previo de las diversas ofertas de trabajo antes de decidir donde aplicar; sin embargo, una vez que un anuncio ha sido analizado y rechazado, la probabilidad de volver a este es menor que la de aplicar a una nueva oferta.

Una particularidad del microfundamento “*stock-flow*” reside en la distinción entre el término de stock, referente al número de desempleados y vacantes al comienzo del intervalo temporal analizado (semana, mes, trimestre, etc.), y el concepto de flujo, que hace referencia a los nuevos desempleados y vacantes que surgen durante ese intervalo de tiempo. Por lo que, los desempleados y vacantes que forman parte del stock al inicio del periodo  $t$ , no se emparejarán debido a que no lo hicieron en el periodo anterior  $t - 1$ . Por tanto, el stock de desempleados  $U$  existente al principio de un periodo se emparejará únicamente con el flujo de nuevas vacantes  $vf$ ; de similar forma sucederá para el stock de puestos de trabajos publicados  $V$ .

Coles y Smith consideran un periodo de tiempo de longitud infinitesimal e ignoran la posibilidad de un emparejamiento entre el flujo de nuevos desempleados y el correspondiente a las nuevas vacantes. De modo que, la probabilidad de que una nueva vacante sea emparejada a su entrada en el mercado es  $1 - \alpha^U$ , siendo  $v(1 - \alpha^U)$  el número de emparejamientos formados a partir de las nuevas vacantes generadas; la misma sistemática se emplea en el caso de los nuevos desempleados, obteniéndose una función de emparejamiento tal que así:

$$M = v(1 - \alpha^U) + u(1 - \alpha^V) \quad 0 < \alpha < 1 \quad (2.7)$$

Teniendo en cuenta la ecuación (2.7), la tasa de emparejamiento para el stock de desempleados resulta ser  $v(1 - \alpha^U)/U$ ; para los nuevos trabajadores  $(1 - \alpha^V)$ .

### 2.3.6. Agregación sobre mercados diferentes.

En este último fundamento, se discute sobre el desequilibrio en los micro mercados y la movilidad laboral limitada. Se considera el supuesto de que la economía está dividida en micro mercados (islas laborales) que no sufren fricciones en su interior, pero sufren de un desequilibrio en el sentido de que la demanda de trabajo en cada mercado no es igual a la oferta. Además, no hay movilidad laboral entre mercados. Esta suposición de ausencia de conexión puede ser interpretada como una fuente de fricción que proporciona fundamento a la existencia de una función de emparejamiento. Esto implica que los mercados con desempleo (oferta de trabajo > demanda de trabajo) pueden coexistir con mercados que tienen vacantes de trabajo (oferta de trabajo < demanda de trabajo), no teniendo ningún mercado simultáneamente ofertas y demandas de trabajo. La agregación de todos estos mercados proporciona una función agregada que contiene ambos elementos, vacantes y desempleados. Con una perfecta movilidad de los trabajadores entre “islas” no existiría la función de emparejamiento agregada.

## 2.4. Problemas de agregación.

### 2.4.1. Problemas de agregación temporal.

El uso de datos discretos en el tiempo puede generar problemas de agregación temporal si solo se consideran en la función de emparejamiento los stocks de ciertas variables (desempleados y vacantes en nuestro caso) y no se tienen en cuenta los flujos de entrada a dichos stocks durante el periodo de tiempo analizado. Por tanto, para lograr una mayor aproximación a la realidad, debemos añadir a la función de emparejamiento descrita anteriormente los flujos de nuevos desempleados y de nuevas vacantes en cada periodo  $t$ :

$$M_t = AU_t^{\alpha_1}uf_t^{\alpha_2}V_t^{\alpha_3}vf_t^{\alpha_4} + \varepsilon_t \quad 0 < \alpha_1, \alpha_2, \alpha_3, \alpha_4 < 1 \quad (2.8)$$

La ecuación (2.8) muestra los emparejamientos producidos en un periodo de tiempo (meses, cuatrimestres, años, etc), de ahí la nomenclatura empleada, incluyendo el subíndice  $t$  en cada una de las variables explicativas. El término  $A$  hace referencia al factor tecnológico procedente de la función de producción de Cobb-Douglas;  $U$  y  $V$  expresan los stocks de desempleados y vacantes al comienzo del periodo, respectivamente; el flujo de nuevos desempleados que surgen durante el rango temporal considerado viene definido por la variable  $uf$  y en el caso de las nuevas vacantes,  $vf$ . Los exponentes  $\alpha_1$ ,  $\alpha_2$ ,  $\alpha_3$  y  $\alpha_4$  son términos de elasticidad, los cuales muestran la variación del flujo de colocaciones al variar

una de las variables descritas  $U$ ,  $V$ ,  $uf$  o  $vf$ , manteniéndose el resto constante, ya que los modelos econométricos aplican el concepto *ceteris paribus*. Por último, la variable  $\varepsilon_t$  representa el término de perturbación aleatoria.

Burdett *et al.* (1994) demuestran que el uso único de los stocks iniciales al comienzo de un periodo de tiempo  $t - 1$ , tal y como se especifica en la ecuación (2.2), genera un sesgo en las elasticidades resultantes de la función de emparejamiento.

#### **2.4.2. Problemas de agregación espacial.**

El otro problema que vincula la teoría de la producción y la tecnología del emparejamiento laboral es la agregación a través del espacio (ya sea geográfico u ocupacional). Al igual que en la literatura empírica sobre la función de producción, la mayoría de los autores que han estimado funciones de emparejamiento han agregado espacialmente el número de trabajadores desempleados y de vacantes y han usado los datos agregados para explicar el flujo de colocaciones laborales en todo el espacio considerado (toda la zona geográfica, por ejemplo). Esta práctica trata a toda la economía agregada como un mercado laboral único, ignorando el hecho de que podría tratarse de una colección de mercados laborales espacialmente distintos con posiblemente poca interacción. La cuestión relevante es si la agregación de datos procedentes de mercados laborales locales separados acaba sesgando las estimaciones resultantes de la función de emparejamiento.

Coles y Smith (1996) argumentan que la agregación espacial podría sesgar los resultados de la función de emparejamiento hacia rendimientos constantes a escala, cuando en realidad podrían ser crecientes. La intuición subyacente a este planteamiento es que, si replicásemos un mercado laboral de un tamaño determinado y con un número determinado de buscadores, se duplicaría el número de colocaciones si no hubiese interacción entre los dos mercados –los rendimientos de escala serían constantes, por tanto. Pero si hay interacción entre los dos mercados, el número de colocaciones debería aumentar más del doble, porque ahora se pueden formar más colocaciones cruzadas (de los buscadores de un mercado con los buscadores del otro). Entonces, con las interacciones entre mercados, las colocaciones más que se duplican cuando el número de buscadores se duplica, lo que implica rendimientos crecientes a escala. Dado que es probable que las interacciones sean más comunes en mercados más densos, Coles y Smith concluyen que, en la estimación de la función de emparejamiento, la densidad del mercado puede ser más importante que el propio tamaño del mercado, algo para lo que encuentran evidencia en su estudio. De hecho, en su estudio del mercado laboral (en Inglaterra y Gales), encuentran rendimientos constantes a



escala en promedio, pero con los mercados más densos ofreciendo tasas de emparejamiento más altas para un tamaño dado de los grupos de vacantes y desempleo.

### **3. METODOLOGÍA ECONOMÉTRICA: DATOS DE PANEL.**

#### **3.1. Introducción a la metodología Datos de Panel.**

La metodología de Datos de Panel es una forma de organización o agrupación de un conjunto de datos estadísticos, empleada, fundamentalmente, en los campos de microeconomía, macroeconomía y finanzas, en la cual se analizan las mismas unidades muestrales (que pueden ser individuos, empresas, países, etc.) en diferentes periodos de tiempo. Dentro de la metodología descrita, se diferencian dos tipos de datos de panel:

- Datos de panel balanceado: todas las unidades muestrales tienen observaciones en todos los periodos de tiempo estudiados.
- Datos de panel no balanceado: no todas las unidades muestrales son observadas en todos los periodos de tiempo estudiados.

Esta metodología es una derivación de la técnica agrupación de cortes transversales de datos correspondientes a diferentes momentos temporales (en inglés *pooled cross-sectional time-series data*), donde las unidades analizadas en un instante de tiempo no tienen por qué coincidir con las del periodo siguiente, –por ejemplo, se realiza una encuesta para las próximas elecciones generales, durante 3 semanas, y en cada una de ellas se seleccionan 200 encuestados de forma aleatoria, aquellos que han sido encuestados una semana determinada, probablemente, no lo serán en la siguiente.

Los Datos de Panel presentan una serie de características, las cuales hacen de esta una metodología atractiva en el ámbito de la economía. Entre ellas, se destacan las siguientes:

- 1) Realiza un seguimiento de las unidades muestrales en numerosos puntos temporales equidistantes (meses, trimestres, años), mientras que en el modelo de datos temporales de corte transversal se consideran un par de instantes en el tiempo, no necesariamente equidistantes.
- 2) El panel de datos permite controlar no solo la heterogeneidad observada (a través de las variables explicativas consideradas en el modelo) sino también la heterogeneidad inobservada que pueda existir en las diferentes unidades muestrales que componen el panel. En las muestras de datos puede haber factores explicativos no medidos o no contemplados, los cuales afectan al comportamiento de los agentes económicos analizados (empresas, países, etc.). Si estas variables relevantes son omitidas del modelo estimado, se produce un sesgo en la estimación –véase Gujarati y Porter (2010). La misma situación acontece con las variables de series de tiempo que influyen por igual en el comportamiento de todas las unidades,

pero dicho efecto es diferente para cada uno de los periodos considerados. La técnica de datos de panel permite corregir la problemática descrita mediante estimación de un “efecto individual” para cada unidad del panel.

- 3) Proporcionan una mayor cantidad de información sobre la variabilidad de la variable endógena al combinar series de tiempo con observaciones de corte transversal. De esta forma se tienen dos fuentes de variabilidad, la propia de comparar diferentes unidades en un periodo dado y la propia evolución de las diferentes unidades en el tiempo; esta mayor disponibilidad de información permite mitigar posibles problemas de multicolinealidad –Kennedy (2008).
- 4) Los datos de series de tiempo y de sección transversal presentan limitaciones, cuando se emplean por separado, a la hora de resolver determinados problemas—por ejemplo, el problema de separación de la existencia de economías de escala (fenómeno transversal) frente al cambio tecnológico (fenómeno temporal) en el análisis de las funciones de producción. En el caso de datos de corte transversal, se pueden comparar los costes de medianas y grandes empresas, pero no es posible estimar los efectos del cambio tecnológico, ya que los datos provienen únicamente de un momento en el tiempo. En contrapartida, si consideramos los datos de series de tiempo de una única empresa, no se puede identificar si los cambios en los costes de producción son producidos por el cambio tecnológico o por una modificación en las dimensiones de la empresa. Empleando los datos de panel, se evita la situación descrita, ya que se analiza una muestra común en diversos periodos de tiempo.
- 5) Permite un análisis óptimo del ajuste dinámico. El conocimiento de las reacciones dinámicas de los individuos es un factor crucial para entender el fenómeno económico estudiado.

### **3.2. Modelos para la estimación de Datos de Panel.**

En los siguientes subapartados, se analizan las cuatro especificaciones habituales de un panel de datos; si bien, en nuestro Trabajo de Fin de Máster nos centraremos en dos de ellas, la especificación de efectos fijos y la de efectos aleatorios:

- Modelo de datos de panel agrupados o, simplemente, regresión agrupada.
- Modelo de efectos fijos o intra-grupos (*within estimation*).
- Modelo entre-grupos (*between estimation*).
- Modelo de efectos aleatorios.

### 3.2.1. Modelo de datos de panel agrupados.

El enfoque más simple para estimar los datos de panel es mediante la regresión agrupada, es decir, omitiendo las dimensiones espaciales y temporales de los datos agrupados –se supone que todos los datos se refieren a un mismo periodo temporal cuando se realiza la estimación, considerando una “gran” regresión. El modelo viene expresado por la ecuación:

$$y_{it} = \beta_0 + x_{it}\beta_k + \varepsilon_{it} \quad (3.1)$$

La expresión (3.1) se caracteriza por obviar las series de tiempo y la naturaleza de corte transversal. Además, el intercepto permanece constante para todas las unidades muestrales, a diferencia del modelo de efectos fijos o el de efectos aleatorios, donde cada individuo toma un valor característico. Los vectores  $x_i$  y  $\beta_k$  representan las variables explicativas del modelo y los coeficientes de estas variables (pendientes), respectivamente. Por último, el vector  $\varepsilon_{it}$  se denomina perturbación estocástica.

Una particularidad del modelo de regresión agrupada y al mismo tiempo, la principal problemática, es que presenta valores muy bajos del estadístico Durbin-Watson, lo que muestra indicios de autocorrelación –véase Gujarati y Porter (2010).

### 3.2.2. Modelo de efectos fijos o estimación intra-grupos (*within estimation*).

Para introducir el modelo de efectos fijos, es necesario partir de una representación (lineal) del modelo de panel más genérica. En este modelo, todas las unidades muestrales tienen la misma pendiente, pero diferente punto de intersección –es decir, la heterogeneidad inobservable se introduce en el intercepto del modelo. Simplificando y considerando únicamente una observación, se obtiene la expresión:

$$y_{it} = \beta_0 + x_{it}\beta + z_i\delta + u_i + \varepsilon_{it} \quad (3.2)$$

donde  $x_{it}$  es un vector de  $k$  variables que varían en función del individuo  $i$  y del tiempo  $t$ ,  $\beta$  es el vector de coeficientes del vector  $x_{it}$ ,  $z_i$  es un vector que contempla las  $l$  variables explicativas que varían por individuo pero resultan invariantes en el tiempo,  $\delta$  el vector de coeficientes de  $z$ ,  $u_i$  expresa los efectos individuales (se suponen estables en el tiempo) y  $\varepsilon_{it}$  es el término de error correspondiente a cada observación disponible.

El modelo de efectos fijos se emplea cuando existe correlación entre el vector de efectos individuales y algunos de los regresores o variables explicativas –el test de Hausman

permite contrastar la existencia de dicha correlación. En caso contrario, en ausencia de correlación entre  $x_{it}$  y  $u_i$ , se demuestra que el modelo de efectos aleatorios es consistente y más eficiente (Baum, 2006) –este modelo se comenta en el siguiente subapartado.

El modelo de efectos fijos se puede estimar de dos formas. Por un lado, mediante la inclusión de variables “*dummies*” para cada individuo (omitiendo la intersección). El problema de esta estimación es que hay que generar tantas *dummies* como unidades muestrales tenga el panel, con la correspondiente pérdida de grados de libertad en la estimación (dado un tamaño muestral determinado). La segunda estrategia para estimar el modelo de efectos fijos pasa por la eliminación del término que contiene los efectos individuales, mediante la transformación que se expresa a continuación –aunque los efectos individuales se eliminan de la ecuación, se sigue pudiendo estimar de forma consistente los parámetros (pendientes)  $\beta$  asociados a las variables explicativas que muestran variabilidad temporal. Esta transformación considera, para cada variable del modelo, los valores promedios de cada unidad muestral durante el periodo de tiempo estudiado, es decir,  $\bar{y}_i = \frac{1}{T} \sum_{t=1}^T y_{it}$ ,  $\bar{x}_i = \frac{1}{T} \sum_{t=1}^T x_{it}$ ,  $\bar{\varepsilon}_i = \frac{1}{T} \sum_{t=1}^T \varepsilon_{it}$ . Aplicando la siguiente transformación en la ecuación (3.2), se obtiene:

$$y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)\beta + (z_i - \bar{z}_i)\delta + u_i - \bar{u}_i + (\varepsilon_{it} - \bar{\varepsilon}_i) \quad (3.2)$$

que implica

$$\tilde{y}_{it} = \beta_0 + \tilde{x}_{it}\beta + \tilde{\varepsilon}_{it} \quad (3.3)$$

Como se observa, el término de efectos individuales y las restantes variables que no cambian en el tiempo ( $z_i$ ) –como pueden ser el género, la raza o la religión de las personas– desaparecen de la ecuación; por ejemplo, si la variable  $z_1$  de la unidad  $i$ -ésima del panel toma todos los periodos el valor 100, la media para el periodo completo de  $z_1$  será 100 y  $z_1 - \bar{z}_1$  será igual a cero en todos los periodos. Este modelo posee una fuerza explicativa, tiene sentido estimarlo, si para cada individuo del panel las desviaciones de la variable y respecto a su media está significativamente correlacionada con las desviaciones de la variable  $x$  respecto a su media. De ahí su denominación “*within estimator*”, ya que depende de la variación temporal que se observa dentro de las diferentes unidades muestrales.

### 3.2.3. Modelo entre-grupos (*between estimation*).

El modelo anterior utiliza la variación que se produce de una observación a la siguiente dentro de cada unidad muestral para explicar los determinantes de la variabilidad de la variable explicada. Otra alternativa para explicar dicha variable consiste en ignorar las variaciones intra-unidad anteriores y centrarse en las variaciones observadas cuando nos movemos de unas unidades muestrales a otras, dando lugar a la llamada estimación entre-grupos. Para implementar este modelo, se calculan los valores medios de las variables del modelo para cada unidad del panel y se estima la siguiente ecuación:

$$\bar{y}_i = \beta_0 + \bar{x}_i\beta + \bar{z}_i\delta + u_i + \bar{\varepsilon}_i \quad (3.4)$$

donde el vector  $\bar{x}_i$  expresa el valor promedio de las variables explicativas por individuo,  $\bar{z}_i$  es un vector que contiene el valor promedio de las variables explicativas que varían por individuo, pero permanecen invariantes en el tiempo,  $\bar{u}_i$  muestra los efectos individuales de cada unidad muestral y  $\bar{\varepsilon}_i$  expresa el valor promedio del término de perturbación aleatoria por cada individuo.

Esta ecuación muestra que si existe correlación entre el término  $u_i (= \bar{u}_i)$  y alguna variable explicativa en el modelo, el supuesto media condicional cero no se mantiene y dicho estimador origina resultados inconsistentes (Baum 2006).

### 3.2.4. Modelo de efectos aleatorios.

Si no existe correlación entre el vector de efectos individuales y las variables explicativas, entonces, los datos de panel se pueden estimar mediante el modelo de efectos aleatorios de forma más eficiente que mediante el modelo de efectos fijos. Al igual que el modelo de efectos fijos, se admite un intercepto diferente para cada individuo, pero dicho efecto individual se estima a través de un procedimiento diferente.

El modelo de efectos fijos tiene dos inconvenientes principales:

1. Al incluir variables ficticias, si son muchas, perdemos grados de libertad. Si pudiéramos encontrar alguna forma de evitar esta pérdida, podríamos producir una estimación más eficiente de las pendientes del modelo.
2. La transformación “*within*” elimina todas las variables explicativas que no varían dentro de cada individuo. Esto significa que cualquier variable explicativa que es invariante en el tiempo, desaparece, por lo que no podemos estimar su correspondiente coeficiente.

La segunda forma de permitir efectos individuales (diferentes interceptos), se basa en el modelo de efectos aleatorios, diseñado para superar estos dos inconvenientes del modelo de efectos fijos. Este modelo es similar al modelo de efectos fijos en que postula un término constante diferente para cada individuo, pero interpreta estas constantes diferentes de una manera diferente. El modelo considera que las diferentes constantes siguen una determinada distribución de probabilidad, normalmente una Normal de media cero y varianza  $\sigma_u^2$ . De esta forma, el modelo a estimar quedaría expresado de la siguiente forma:

$$y_{it} = \beta_0 + x_{it}\beta + z_i\delta + (u_i + \epsilon_{it}) \quad (3.5)$$

$$u_i \sim N(0, \sigma_u^2), \quad \epsilon_{it} \sim N(0, \sigma_\epsilon^2), \quad E(\epsilon_{it}|x_{it}) = 0, \quad E(u_i|x_{it}) = 0$$

Como se puede observar, tenemos una especificación econométrica en la que hay una constante general, un conjunto de variables explicativas (con sus coeficientes de interés) y un término de error compuesto. Este error compuesto tiene dos partes. Para cada individuo, una parte es el efecto fijo  $u_i$ , que mide el grado en que la constante del individuo difiere de la constante general del modelo. La otra parte es solo el error aleatorio tradicional  $\epsilon_{it}$  que explica la existencia de perturbaciones aleatorias que pueden afectar al individuo en cada periodo tiempo. Para un individuo en particular  $i$ ,  $u_i$  toma el mismo valor en todos los períodos de tiempo, mientras que  $\epsilon_{it}$  es diferente en cada período de tiempo.

El “truco” para para estimar por efectos aleatorios es reconocer que la matriz de varianza-covarianza de este error compuesto no es esférica (es decir, no todos los elementos de fuera de la diagonal de esta matriz son cero): para todas las observaciones de un individuo específico,  $u_i$  vale lo mismo, por lo que los errores compuestos de todas esas observaciones están correlacionados –se suele suponer que las observaciones correspondientes a diferentes individuos tienen una correlación cero entre sus errores compuestos. Este patrón de correlaciones crea una matriz de varianza-covarianza con un patrón especial que puede ser estimada usando Mínimos Cuadrados Generalizado, método que realiza una transformación de los datos que da lugar a una matriz de varianza-covarianza esférica que puede, ahora sí, ser estimada por Mínimos Cuadrados Ordinarios

## **4. DESCRIPCIÓN DE LOS DATOS EMPLEADOS SOBRE EL EMPAREJAMIENTO LABORAL EN ESPAÑA A NIVEL REGIONAL.**

### **4.1. Análisis descriptivo de las variables laborales. Series trimestrales. Período 2006-2018.**

En el presente apartado se analizarán las distintas variables laborales que intervienen en el proceso de emparejamiento, permitiendo además explorar el grado de eficiencia del emparejamiento laboral en España a nivel de Comunidades Autónomas. Como se ha discutido en secciones anteriores, la función de emparejamiento explica el comportamiento y la relación existente entre la generación de empleo (colocaciones o contratos iniciales) y los stocks iniciales y flujos de entradas de desempleados y vacantes.

Las variables que se van a analizar, durante el periodo 2006-2018 (series trimestrales), se indican a continuación.

- Stock inicial de desempleados ( $U$ ).
- Stock inicial de vacantes ( $V$ ).
- Flujo de nuevos desempleados ( $uf$ ).
- Flujo de nuevas vacantes ( $vf$ ).
- Colocaciones o contratos iniciales ( $M$ ).

Para la obtención de los valores de las variables anteriores en el periodo indicado nos hemos servido de los datos recogidos en la *Estadística de Empleo* que publica el Servicio Público de Empleo Estatal (SEPE) empleando la metodología SISPE (Sistemas de Información de los Servicios Públicos de Empleo). Únicamente en el caso del stock de vacantes se ha recurrido a dos encuestas que registran el dato de vacantes en dos subperiodos distintos: la *Encuesta de Coyuntura Laboral* (ECL), perteneciente al Ministerio de Trabajo, Migraciones y Seguridad Social, que nos proporciona datos del stock de vacantes por CCAA desde el primer trimestre de 2006 hasta el cuarto trimestre de 2012; y la Encuesta Trimestral del Coste Laboral (INE) que publica dicha serie desde el tercer trimestre de 2013 hasta la actualidad. Como se puede apreciar, la serie de stock de vacantes presenta una discontinuidad en los dos primeros trimestres de 2013, por lo tendremos que realizar una intervención en dicha serie para estimar esas observaciones no disponibles.

Afortunadamente, todas estas estadísticas representan las series de datos laborales de forma desagregada a nivel regional, es decir, disponemos de información para cada una de las comunidades autónomas (CCAA) que componen el territorio español, lo cual permite



estimar un panel de datos regional. Las series empleadas se describen en los siguientes apartados.

#### **4.1.1. La Estadística de Empleo del Servicio Público de Empleo Estatal.**

El Servicio Público de Empleo Estatal consiste en un organismo autónomo de la Administración General del Estado, adscrito al Ministerio de Trabajo, Migraciones y Seguridad Social. Este, junto con los Servicios Públicos de Empleo de las Comunidades Autónomas, componen el llamado Sistema Nacional de Empleo. El nombre SEPE sustituye al de INEM (Instituto Nacional de Empleo) desde el año 2003, aunque muchos trabajadores y empresas siguen empleando el término INEM para referirse al SEPE.

El SEPE, a través de su Subdirección General de Estadística e Información, elabora y difunde estadísticas sobre el empleo. Desde el año 2005, dichas estadísticas siguen la metodología SISPE (que sustituye a la metodología SILE, Sistema de Información Laboral de Empleo). Aunque SISPE introduce mejoras significativas respecto a SILE –como, por ejemplo, una mayor uniformidad en la información que las diferentes CCAA transmiten al SEPE–, ninguno de los dos sistemas ha conseguido resolver el problema de generar una serie del stock de vacantes que sea válida para el total de la economía española (no solo para la parte del mercado laboral donde intermedia el SEPE); por lo que en este Trabajo de Fin de Máster realizamos una propuesta de aproximación a dicha serie.

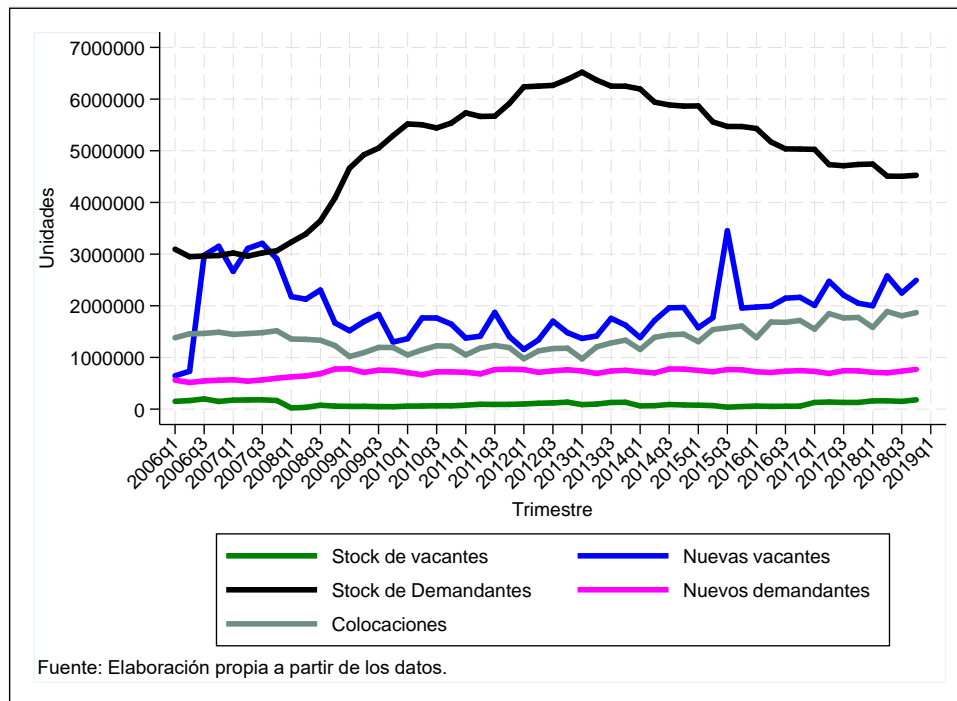
#### **4.2. Evolución anual del mercado de trabajo en España.**

En la figura 4.1 se representa la evolución de las variables laborales analizadas, para el periodo comprendido entre los años 2006 y 2018, las cuales permiten conocer la tendencia que ha seguido el mercado laboral español durante este rango de tiempo, pudiéndose observar, además, los efectos provocados por la crisis económica sufrida en nuestro país a partir del año 2008.

Como se aprecia en la figura, el stock de demandantes ( $U$ ) aumentó considerablemente a partir del año 2008, debido al inicio de la crisis económica, y alcanzó su valor máximo allá por el año 2013, comenzando, de ahí en adelante, un proceso de disminución (el número de desempleados disminuye) vinculado al periodo de recuperación económica que se hizo patente a partir del año 2014. En cuanto al flujo de nuevas vacantes generadas ( $vf$ ), muestra un cierto comportamiento cíclico (procíclico en este caso) y presenta también un comportamiento estacional importante, debido principalmente a que, durante el tercer cuatrimestre de cada año estudiado, su número aumenta significativamente. La justificación

de este pico estacional en el tercer trimestre se debe a que coincide con el periodo estival, en el cual se crean vacantes de forma temporal como consecuencia del fuerte turismo durante el periodo vacacional. Con respecto a las colocaciones ( $M$ ), se observa que presenta una alta correlación con el flujo de vacantes ( $vf$ ), es decir, al aumentar el número de ofertas de trabajo registradas, existe una mayor probabilidad de que los desempleados obtengan una oportunidad laboral, de ahí que ambas variables se comporten de forma sincronizada.

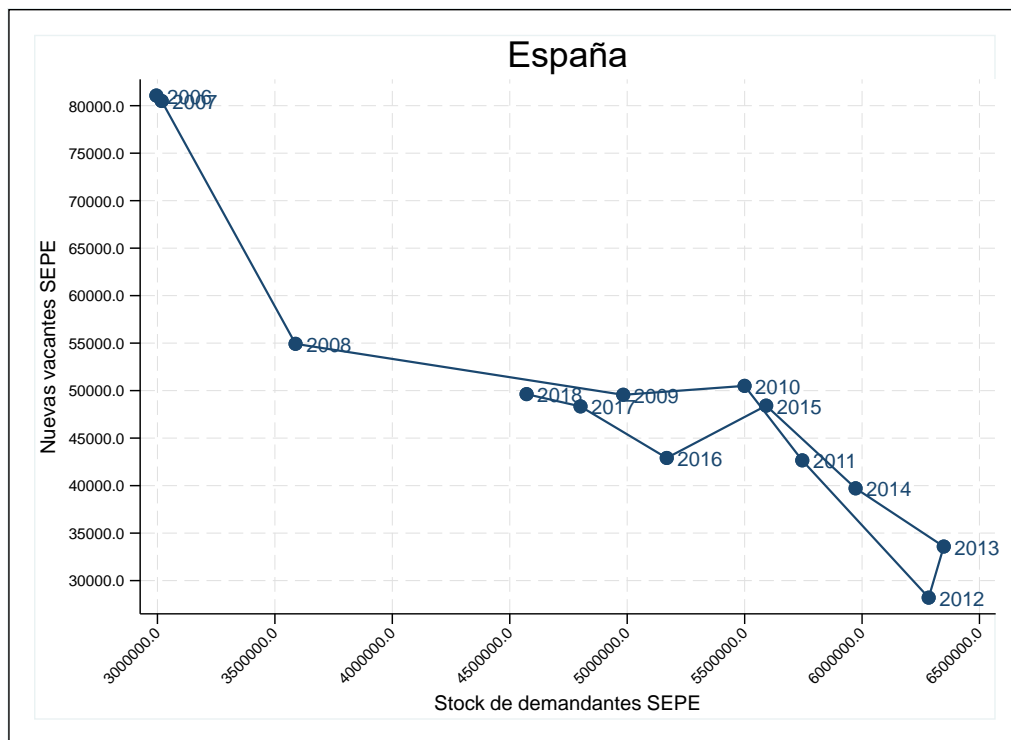
A nuestro entender, durante los primeros años del periodo analizado (2006 a 2008), la información sobre variables laborales no se resulta fiable, ya que se produjo el cambio del sistema de gestión del SEPE, que pasó de SILE a SISPE, con el consiguiente periodo transitorio de ajuste metodológico en las series estadísticas.



**Figura 4.1. Evolución trimestral del mercado de trabajo en España. Periodo 2006:T1–2018:T4.**

Otra forma de observar la evolución o la tendencia del mercado español es a través de la relación empírica vacantes-desempleo (cuyo fundamento teórico viene dado por la relación teórica denominada curva de Beveridge). Aunque esta relación se suele representar con los stocks de vacantes y desempleo, nosotros la vamos a representar como una relación empírica entre el stock de demandantes ( $U$ ) y el flujo de nuevas vacantes ( $vf$ ); dado que existe evidencia a nivel español de que el emparejamiento más frecuente es de tipo stock-flujo –véase, por ejemplo, Álvarez de Toledo *et al.* (2008). En la figura 4.2 se aprecia la situación de recesión sufrida en España durante los años 2008 hasta 2012, observándose una

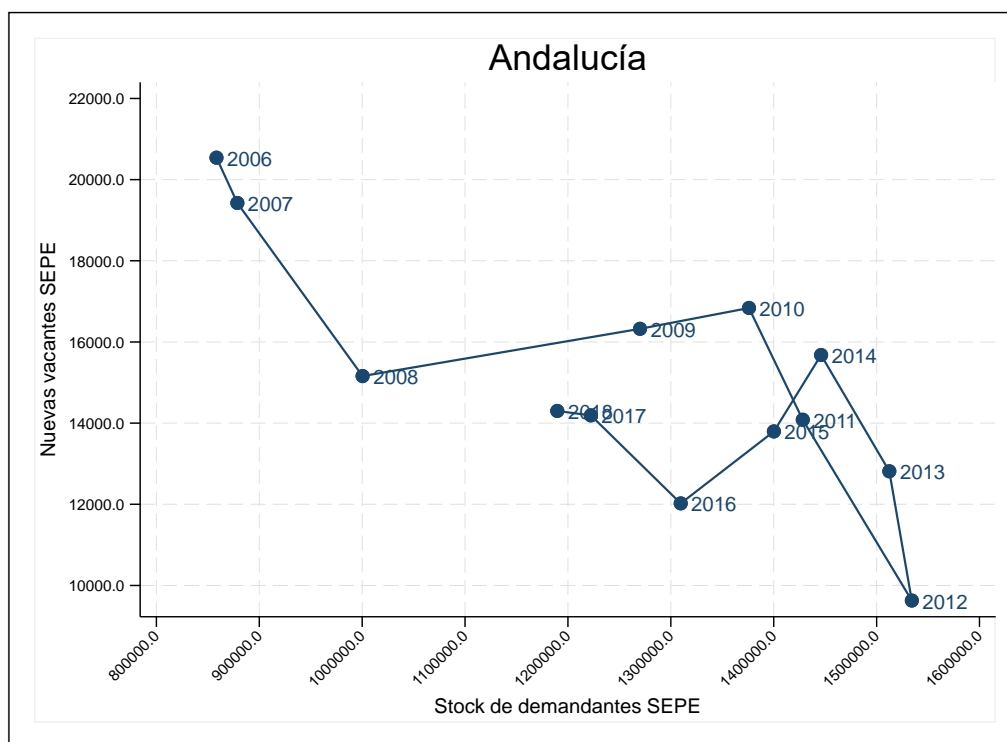
pendiente negativa en la relación, la cual indica que en dicho periodo se produjo un aumento en el número de desempleados (demandantes de empleo) y a su vez, una caída en la generación de nuevos puestos vacantes. Entre 2009 y 2010, se produce una situación que podríamos llamar de falsa recuperación de la economía, y a partir de este último año, sigue agravándose la crisis, alcanzando el stock de demandantes sus valores máximos, entorno a los seis millones y medio de desempleados, en los años 2012–2013. Es a partir de 2014, cuando se constata una recuperación de la actividad económica, ya que el número de desempleados comienza a disminuir, mientras que el flujo de nuevas vacantes se encuentra en pleno crecimiento –el flujo de nuevas vacantes comenzó a aumentar en 2013. Esta tendencia expansiva se sigue apreciando al final del periodo analizado, es decir, 2018.



**Figura 4.2. Curva de Beveridge del mercado de trabajo en España. Periodo 2006:T1–2018:T4.**

Si se realiza el mismo análisis en una de las comunidades autónomas más afectadas por la crisis, Andalucía, se observa un comportamiento similar al del total del país. Desde el año 2006, el número de desempleados crece, pero dicho crecimiento se intensifica a partir de 2008, llegando a su valor máximo en 2012. Entre 2008 y 2010 el flujo de nuevas vacantes aumenta a pesar de que el desempleo sigue creciendo, lo que apunta a una pérdida de eficiencia en el emparejamiento laboral (la curva de Beveridge teórica experimenta un desplazamiento hacia la derecha que la aleja del origen); la comunidad andaluza presenta

una elevada temporalidad del empleo, por lo que cabría esperar que el aumento del flujo de nuevas vacantes tuviera un alto contenido de temporalidad (vinculado a sectores como la agricultura, la construcción o el turismo). En 2013, se frena el ciclo recesivo y la economía andaluza comienza a reducir el volumen de demandantes de empleo y a aumentar su volumen de nuevas vacantes; esta aparente recuperación económica se observa hasta la actualidad.



**Figura 4.3. Curva de Beveridge del mercado de trabajo en Andalucía. Periodo 2006:T1–2018:T4.**

#### 4.3. Evolución trimestral del stock de demandantes por CCAA.

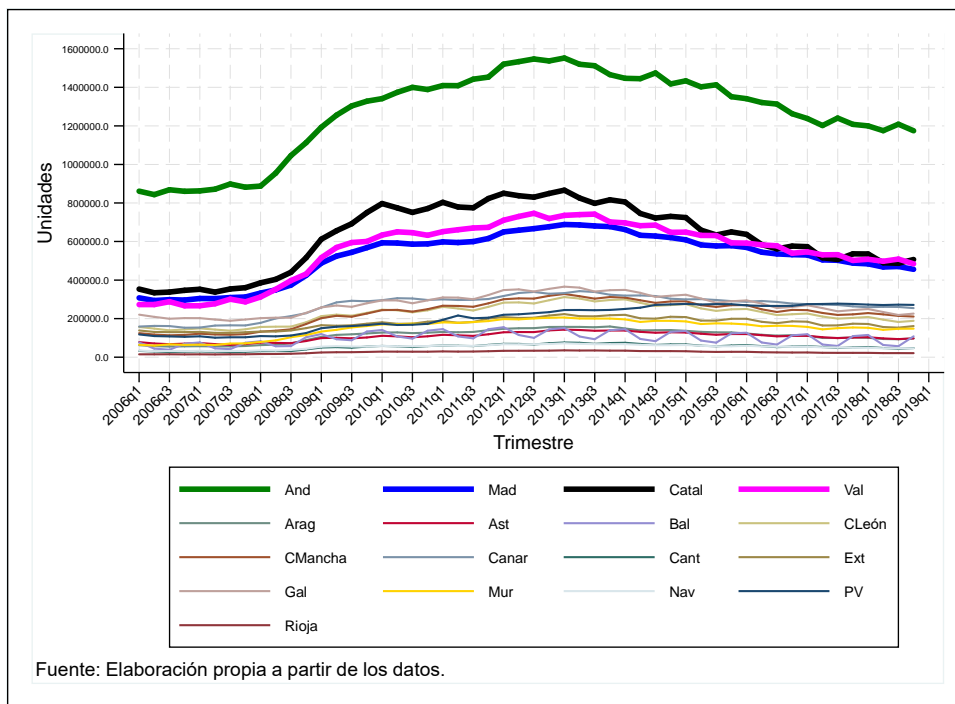
A diferencia de la figura 4.1, la cual refleja la evolución entre 2006 y 2018 de todas las variables laborales influyentes en el emparejamiento laboral para el caso de España, en los siguientes apartados se analizarán dichas variables por separado a nivel autonómico.

Debido a la disparidad de valores en los que se mueven las diferentes comunidades autónomas, se han representado las series en dos escalas diferentes (lineal y logarítmica en base 10); por ejemplo, la comunidad de Cataluña tiene alrededor de ocho millones de habitantes en 2018, y Cantabria, lejos de llegar a las siete cifras, muestra una población en torno a seiscientos mil habitantes.

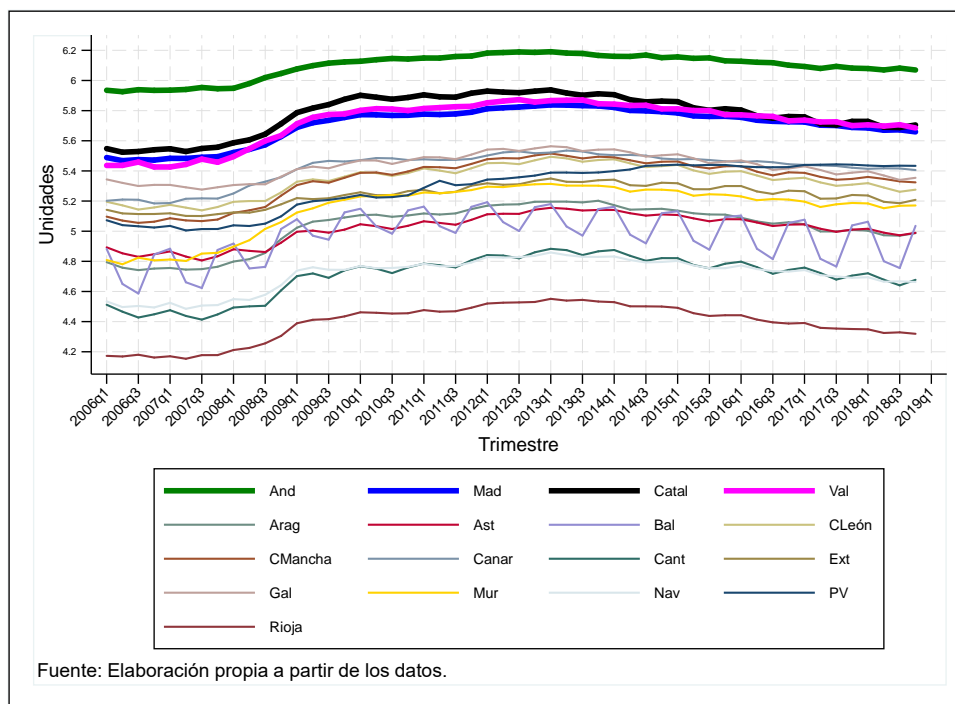
En la figura 4.4 se observa como Andalucía destaca como la comunidad autónoma con mayor número de desempleados durante todo el periodo de tiempo considerado. En 2008, se produce un crecimiento del stock de demandantes en todas las CCAA, siendo 2013 el año

donde se inicia una ligera disminución de la población en desempleo. Por otro lado, Madrid, Cataluña y Valencia, mantienen el mismo comportamiento cíclico que Andalucía, pero en niveles algo inferiores. Concretamente, este comportamiento, observable en el resto de las comunidades a niveles bastante inferiores, se considera anticíclico, ya que se mueve en la dirección contraria que el PIB de la economía española; cuando el PIB cae o se estanca, el desempleo aumenta, y a la inversa.

En lo que se refiere a la estacionalidad intra-anual de la serie, destacamos el caso de la comunidad de Baleares, que presenta un comportamiento donde se aprecia claramente la estacionalidad, concretamente con una caída importante durante el tercer trimestre de cada año, debido al comienzo de la temporada estival, implicando un aumento en la creación de trabajos temporales para cubrir la demanda generada por el turismo y, por consiguiente, una disminución importante del stock de demandantes.



a) Evolución trimestral del stock de demandantes por CCAA en escala lineal. Periodo 2006:T1–2018:T4.



b) Evolución trimestral del stock de demandantes por CCAA en escala logarítmica (base 10).  
Periodo 2006:T1–2018:T4.

**Figura 4.4. Evolución trimestral del stock de demandantes por CCAA. Periodo 2006:T1–2018:T4.**

#### 4.4. Evolución trimestral del stock de vacantes por CCAA.

La serie del stock de vacantes registrados en el SEPE no es representativa del stock total de vacantes de la economía española, ya que la mayor parte de las vacantes que ofertan en el país (más del 95%) no son dadas de alta por el empresario en ninguna oficina pública de empleo. Ante esta falta de representatividad de la serie de vacantes del SEPE, en este Trabajo hemos preferido recurrir a las series de la Encuesta de Coyuntura Laboral (ECL) y a su sucesora, la Encuesta Trimestral del Coste Laboral (ETCL), para medir el nivel de vacantes de la economía. Ambas series realizan un muestreo sobre las empresas y posteriormente, extrapolan la información de las vacantes obtenida al total de la economía. Así, según la metodología de la ECL el ámbito sectorial de la encuesta “se circunscribe a las actividades económicas de la Industria, Construcción y Servicios, excluyéndose de este último sector la Administración Pública, Defensa y Seguridad Social obligatoria, Organismos extraterritoriales y Organizaciones religiosas. Es decir, se corresponde con el sector no Agrario con las exclusiones citadas del sector Servicios.”. Por su parte, la metodología de la ETCL declara que “se investigan las cuentas de cotización cuya actividad económica esté encuadrada en los tres grandes sectores económicos: Industria, Construcción y Servicios, en concreto aquellos centros con actividades económicas comprendidas en las

secciones de la B a la S de la CNAE-09”. Excluye por tanto Agricultura, ganadería, silvicultura y pesca (A), Actividades de los hogares como empleadores de personal doméstico (T) y Actividades de organizaciones y organismos extraterritoriales (U). Por tanto, la ausencia de ciertos sectores, especialmente el de la agricultura, origina la necesidad de un reescalamiento como el expresado en la ecuación (4.1). La idea de este ajuste es usar el flujo de entrada al stock de vacantes del SEPE para reescalar el stock de vacantes que publican ECL y ETCL. En efecto, el ratio del total de nuevas vacantes SEPE en todos los sectores de actividad en relación a las nuevas vacantes SEPE en los sectores correspondientes a aquellos que son contemplados por la ECL-ETCL, indicaría cuánto se incrementa el flujo de nuevas vacantes SEPE al añadir a los sectores contemplados por la ECL-ETCL aquellos sectores excluidos por estas encuestas –por ejemplo, si dicho factor es de 1,1, podríamos concluir que el flujo de nuevas vacantes SEPE correspondiente a los sectores muestreados por la ECL-ETCL se incrementaría un 10% si se incluyeran las nuevas vacantes SEPE correspondientes a los sectores excluidos por dichas encuestas. A partir de aquí, no resulta descabellado suponer que dicho incremento sea similar a nivel del stock de vacantes (ya sea a nivel del stock de vacantes del SEPE, el cual no se publica, o a nivel del stock de vacantes de la ECL-ETCL), ya que el flujo de nuevas vacantes no es más que el flujo de entrada a dicho stock o nivel, por lo que podemos plantear una corrección como la propuesta en la ecuación (4.1) donde se corrige el stock de vacantes con información procedente de su flujo de entrada.

$$V_{total} = V_{ecl-etcl} \cdot \left( \frac{vf}{vf_{Ind} + vf_{Const} + vf_{Servicio\_ECL}} \right) \quad (4.1)$$

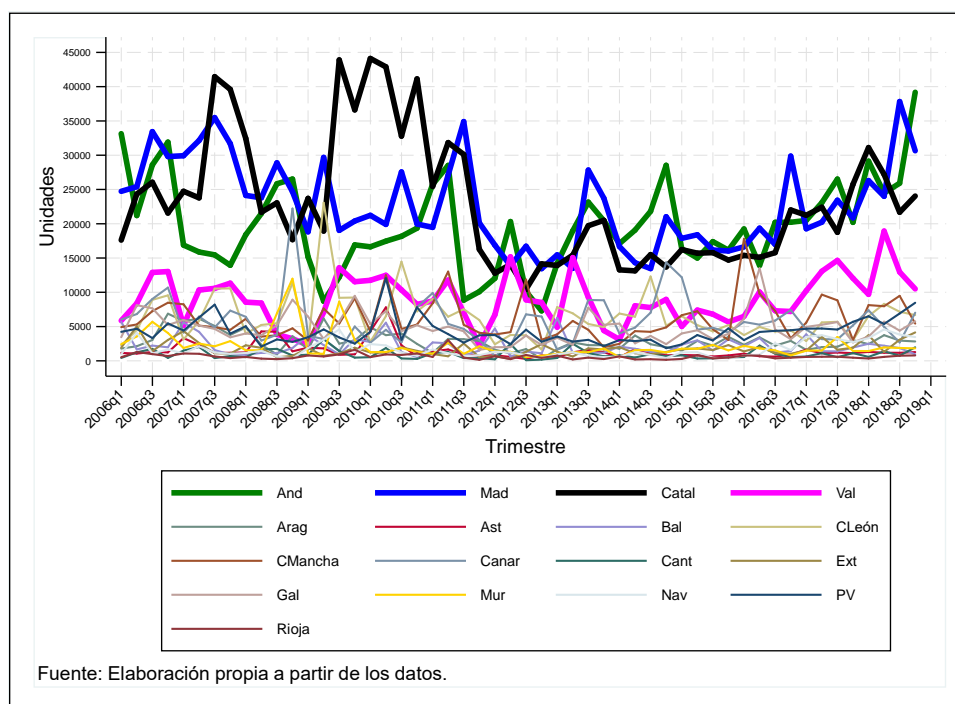
A su vez, como se ha indicado en apartados anteriores, debido a la discontinuidad de los datos de vacantes durante los dos primeros trimestres del año 2013 –a raíz de la transición de la ECL a la ETCL – se ha definido la ecuación (4.2) que permite dar cobertura a este vacío temporal.

$$V_{ecl} = \frac{V_{ecl}[n - 4] + V_{etcl}[n + 4]}{2} \quad (4.2)$$

La expresión anterior se ha empleado para calcular el stock de vacantes en los dos primeros trimestres de 2013, donde no disponemos de información acerca de las ofertas de trabajo publicadas en ese periodo de tiempo. De modo que, para el caso del primer trimestre

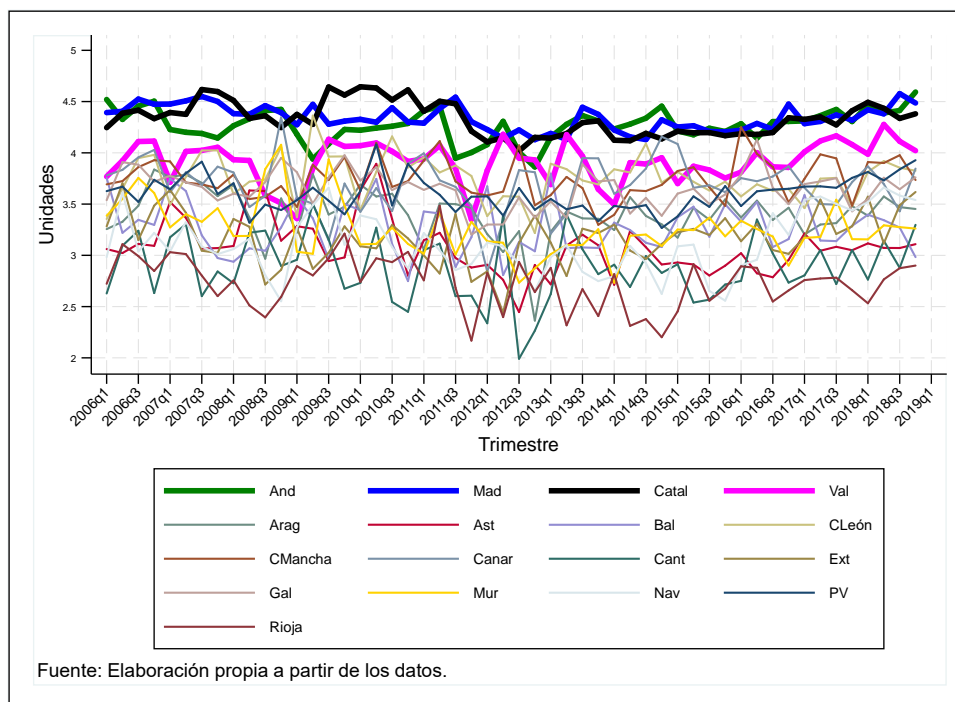
de 2013, el stock de vacantes se expresa como el promedio del primer trimestre en los cuatro años anteriores y posteriores a 2013; de forma similar para el segundo trimestre del mismo año.

Descrito el proceso de obtención de los datos de las vacantes en el mercado laboral español, se han representado los valores obtenidos en la figura 4.5, de la cual se pueden extraer conclusiones interesantes como que las comunidades de Madrid, Cataluña y Andalucía poseen el mayor número de vacantes dentro de nuestro país. El dato no debe extrañar dado que son las regiones con los mercados de trabajo de mayor tamaño. Andalucía basa su volumen de vacantes en sectores tradicionales como la agricultura, la construcción o el turismo, mientras que en Madrid y Barcelona tienen mayor protagonismo relativo el resto de los servicios y el sector industrial. Aunque la volatilidad de la serie dificulta detectar un comportamiento cíclico (o incluso estacional), lo cierto es que la serie muestra un perfil procíclico, cayendo con el PIB hasta el año 2013 y empezando a crecer a un ritmo cada vez mayor desde dicho año, cuando las tasas de crecimiento económico de las regiones españolas se recuperan.



a) Evolución trimestral del stock de vacantes por CCAA en escala lineal. Periodo 2006:T1–2018:T4.





b) Evolución trimestral del stock de vacantes por CCAA en escala logarítmica (base 10). Periodo 2006:T1–2018:T4.

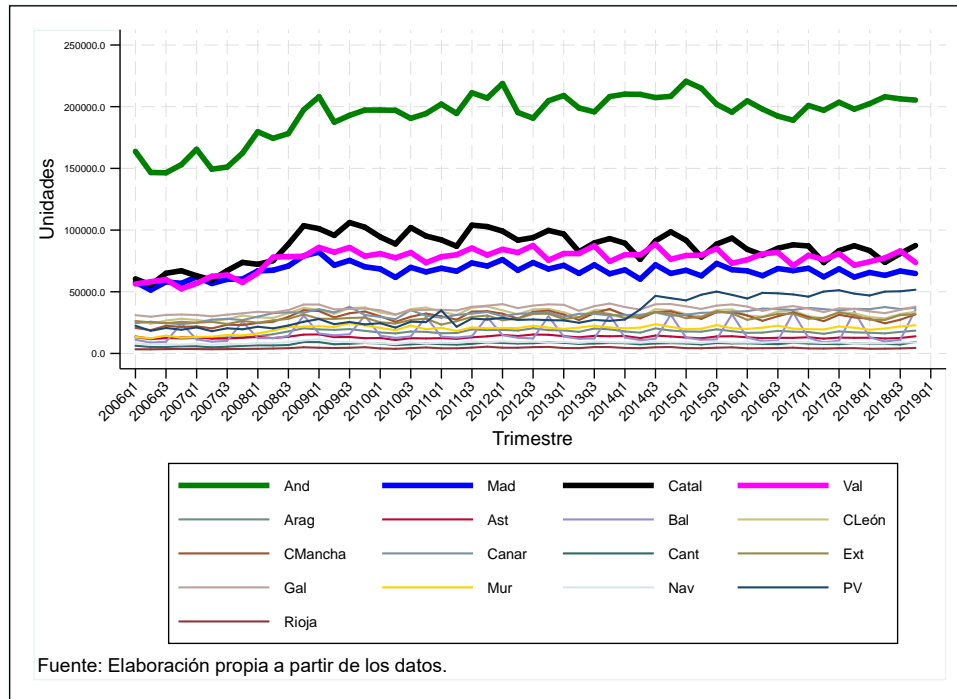
**Figura 4.5. Evolución trimestral del stock de vacantes por CCAA. 2006:T1–2018:T4.**

#### 4.5. Evolución trimestral del flujo de nuevos demandantes por CCAA.

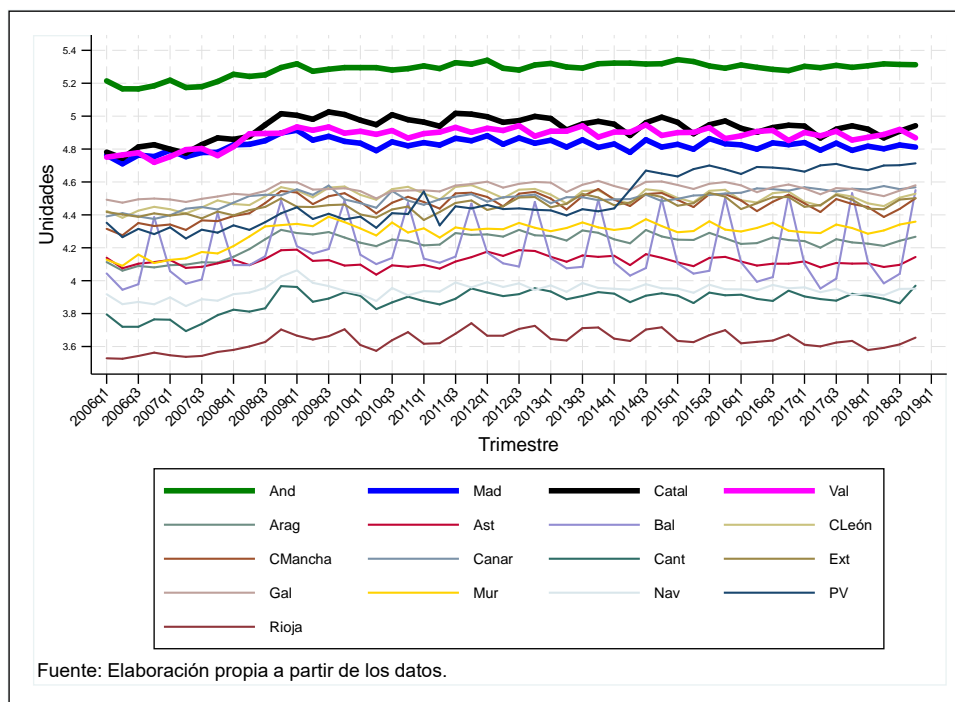
Si analizamos el flujo de nuevos demandantes (figura 4.6), las comunidades de Andalucía, Madrid, Cataluña y Valencia destacan, nuevamente, por ser aquellas donde existe un flujo predominante. Presentando, Andalucía, una gran diferencia con respecto a las demás; hay que tener en cuenta que el trabajo de carácter temporal es un aspecto relevante de esta comunidad, por lo que muchos trabajadores entran al desempleo (y salen del desempleo) multitud de veces dentro del año natural.

El componente anticíclico de la serie es relativamente más suave que el observado en el caso del stock de demandantes, posiblemente porque cuando la economía entra en recuperación/expansión puede haber un efecto “ánimo” que incremente la población activa tanto por el lado del empleo como por el lado del desempleo, por lo que años de expansión no tienen por qué implicar necesariamente caídas importantes en el número de nuevos buscadores de empleo. Por otro lado, el componente estacional se observa claramente en La Rioja y Baleares, siendo en esta última comunidad donde existe una mayor oscilación de la variable; en concreto, el flujo de nuevos buscadores de empleo en las Islas Baleares aumenta considerablemente en el último trimestre de cada año y disminuye en los meses relacionados con la temporada turística de la isla.

En el caso de la Rioja, los picos y valles en el flujo de nuevos demandantes se deben a la época de la vendimia en la zona norte del país, siendo esta coincidente con los meses de septiembre a octubre. De modo que, en este intervalo es cuando se observa un menor número de demandantes, ascendiendo una vez que finaliza la vendimia.



a) Evolución trimestral del flujo de nuevos demandantes por CCAA en escala lineal. Periodo 2006:T1–2018:T4.

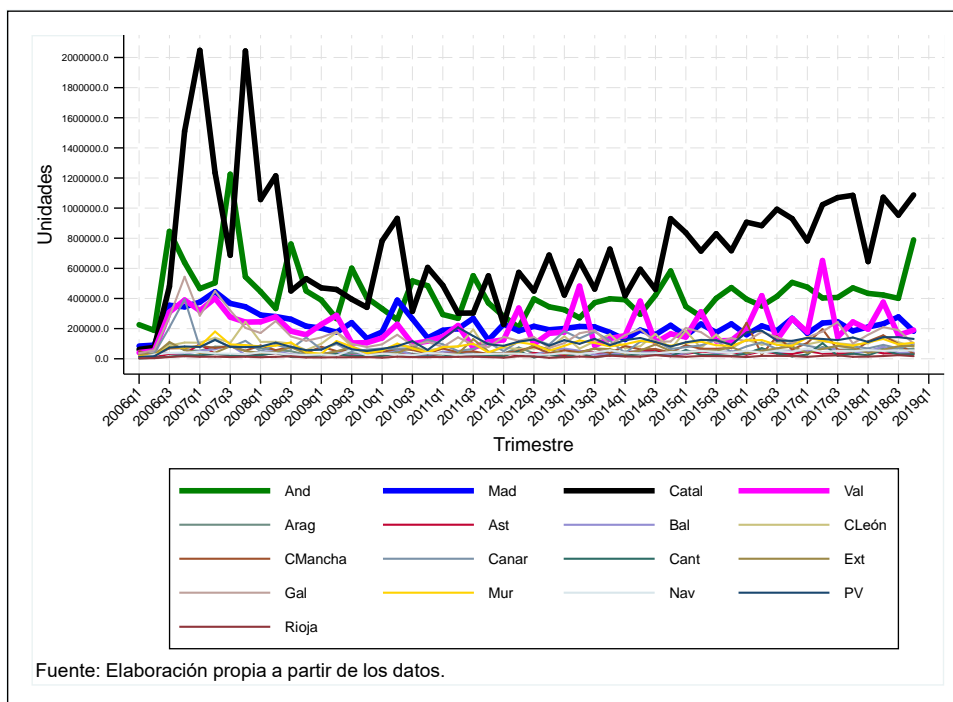


b) Evolución trimestral del flujo de nuevos demandantes por CCAA en escala logarítmica (base 10). Periodo 2006:T1–2018:T4.

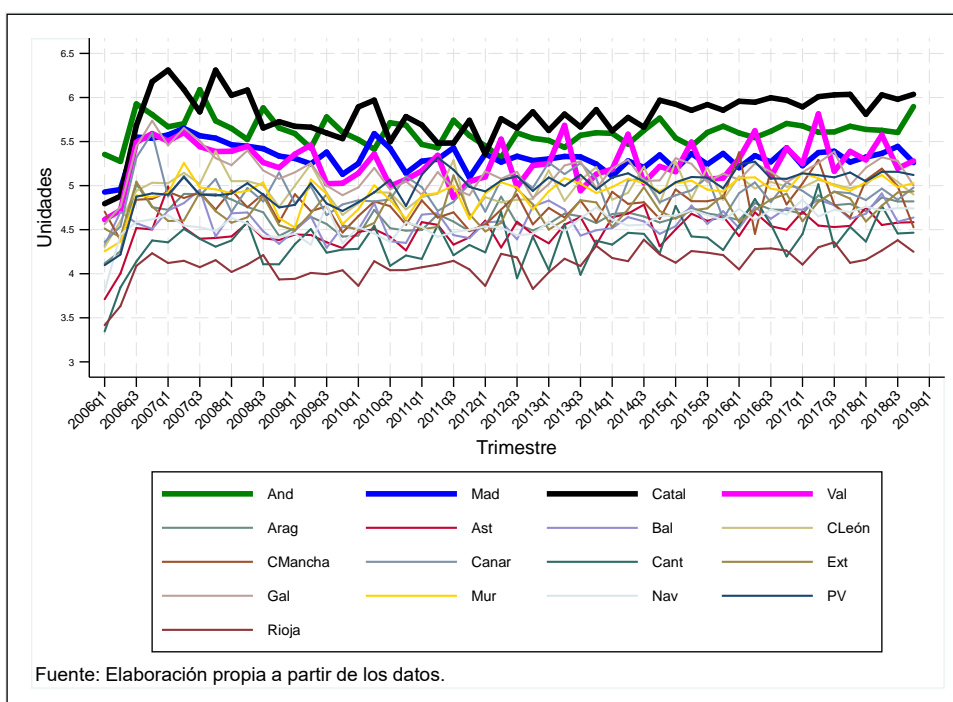
**Figura 4.6. Evolución trimestral del flujo de nuevos demandantes por CCAA. 2006:T1–2018:T4.**

#### 4.6. Evolución trimestral del flujo de nuevas vacantes por CCAA.

Respecto al flujo de nuevas vacantes por CCAA, representado en la figura 4.7, se observa como Cataluña se caracteriza por ser la comunidad que genera, durante el rango temporal objeto de este proyecto, un mayor número de vacantes, de modo similar a lo que sucedía con el stock de vacantes. Esto es debido posiblemente a la importancia de sus sectores servicios e industrial (generadores de empleo); en este último sector (el industrial), destaca, por ejemplo, el de la automoción. A Cataluña, le siguen comunidades como Andalucía, Madrid o Valencia. Casi todas las CCAA presentan volatilidad intra-anual, lo cual es señal de presencia de estacionalidad en la generación de vacantes –véase, por ejemplo, el pico alcista de Valencia en los segundos cuatrimestres de cada año o el de Andalucía en el tercer cuatrimestre.



a) Evolución trimestral del flujo de nuevas vacantes por CCAA en escala lineal. 2006:T1–2018:T4.



b) Evolución trimestral del flujo de nuevas vacantes por CCAA en escala logarítmica (base 10). Periodo 2006:T1–2018:T4.

**Figura 4.7. Evolución trimestral del flujo de nuevas vacantes por CCAA. Periodo 2006:T1–2018:T4.**

#### 4.6.1. Re-escalamiento del flujo de nuevas vacantes por CCAA.

La serie de nuevas vacantes (u ofertas de empleo) que publica el SEPE tiene un problema importante de falta de representatividad estadística si se usa para aproximarse al volumen total de nuevas vacantes de la economía española. Resulta que solo un porcentaje pequeño de las vacantes que se generan en el territorio español se registran en las oficinas públicas de empleo para su gestión y cobertura –la mayoría de las vacantes se cubren (en el sector privado o en el público) sin la ayuda del SEPE. Para estimar este porcentaje, basta con echar un vistazo a las cifras de colocaciones. Cuando una vacante sí registrada en el SEPE se cubre, da lugar a una colocación “con oferta (vacante) previa”. Las colocaciones con oferta previa representan un porcentaje del total de colocaciones inferior al 5%, el porcentaje restante (superior al 95%) se corresponde con las llamadas colocaciones “sin oferta previa”, que son colocaciones donde no se produce ningún registro previo en el SEPE de la vacante que se cubre –en este caso, la empresa se limita a notificar la cobertura de la vacante al SEPE pero sin haber registrado la vacante previamente en el SEPE, el cual, por tanto, no gestiona la búsqueda de candidatos para dicha vacante (la empresa se encarga de la búsqueda de candidatos).

La falta de registro en el SEPE de las vacantes que acaban generando colocaciones sin oferta previa plantea un problema importante a nivel estadístico: los datos del SEPE solo dan información sobre la fecha o el mes de alta de las (pocas) nuevas vacantes que se registran en sus oficinas, y que darán lugar en su momento, cuando se cubran, a colocaciones con oferta previa. Sin embargo, el resto de las vacantes no gestionadas por el SEPE, que son la gran mayoría, también deben ser tenidas en cuenta si se quiere trabajar con el flujo total de vacantes que realmente hay en la economía española.

Por tanto, en base a la problemática descrita surge la necesidad de realizar un re-escalamiento del flujo de nuevas vacantes con la finalidad de obtener una serie de datos que sea más representativa del volumen total de nuevas vacantes que existe en la economía española. En este Trabajo de Fin de Máster vamos a usar una variante del re-escalamiento propuesto por Antolín (1994), y que viene representado por la ecuación (4.3) –Antolín (1994) propone el re-escalamiento a nivel del stock de vacantes, mientras que aquí se está aplicando al flujo de nuevas vacantes.

$$\frac{vf_{no\ sepe}}{vf_{sepe}} = k \frac{M_{sinofprev}}{M_{ofprev}} \Rightarrow$$

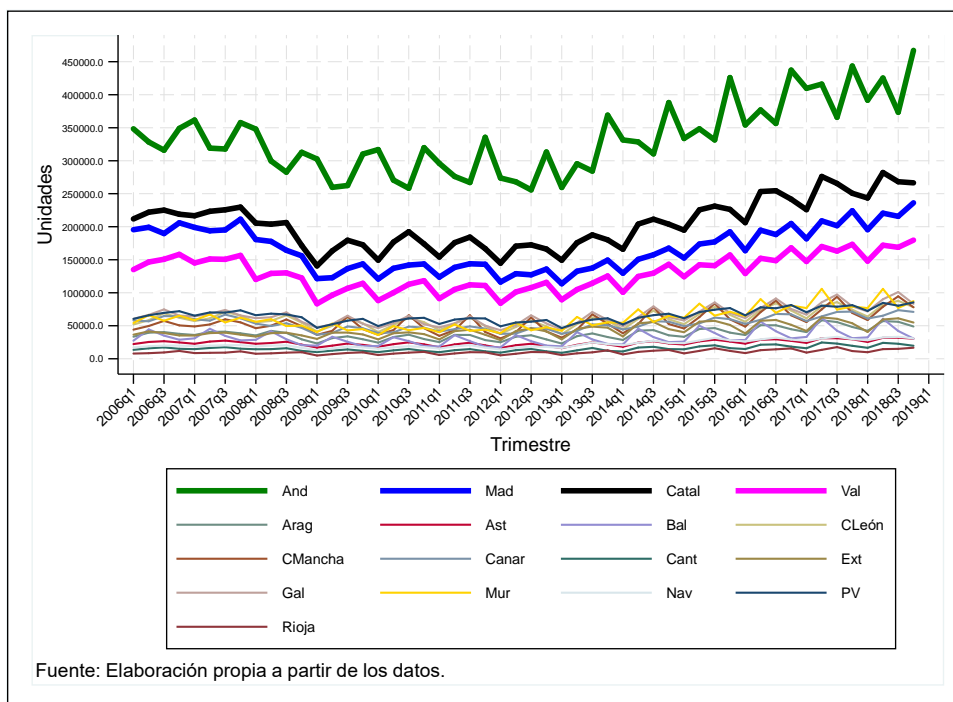
$$vf_{total} = vf_{no\ sepe} + vf_{sepe} = \left(1 + k \frac{M_{sinofprev}}{M_{ofprev}}\right) \cdot vf_{sepe} \quad (4.3)$$

Donde  $vf_{total}$  representa la estimación mensual del flujo total de nuevas vacantes,  $vf_{sepe}$  (o simplemente  $vf$ ) representa el flujo mensual de nuevas vacantes que se registra en las oficinas de empleo del SEPE,  $M_{sinofprev}$  y  $M_{ofprev}$  hacen referencia a las colocaciones sin oferta y con oferta previa, respectivamente, y  $k$  es un factor que expresa la eficiencia relativa del sector privado (frente al SEPE) en la cobertura de vacantes. Aunque Antolín propone un factor  $k$  de 0,25 (según este autor, a principios de los noventa en España, el sector privado era 4 veces más rápido que el INEM-SEPE cubriendo una vacante), nosotros vamos a emplear un factor de  $k = 1$ ; para tomar esta decisión nos basamos en los resultados del trabajo de Álvarez de Toledo *et al.* (2008) y en el hecho de que las mejoras informáticas en la intermediación laboral pública han aumentado la eficiencia relativa del SEPE en su desempeño de dar cobertura a las vacantes que le llegan.

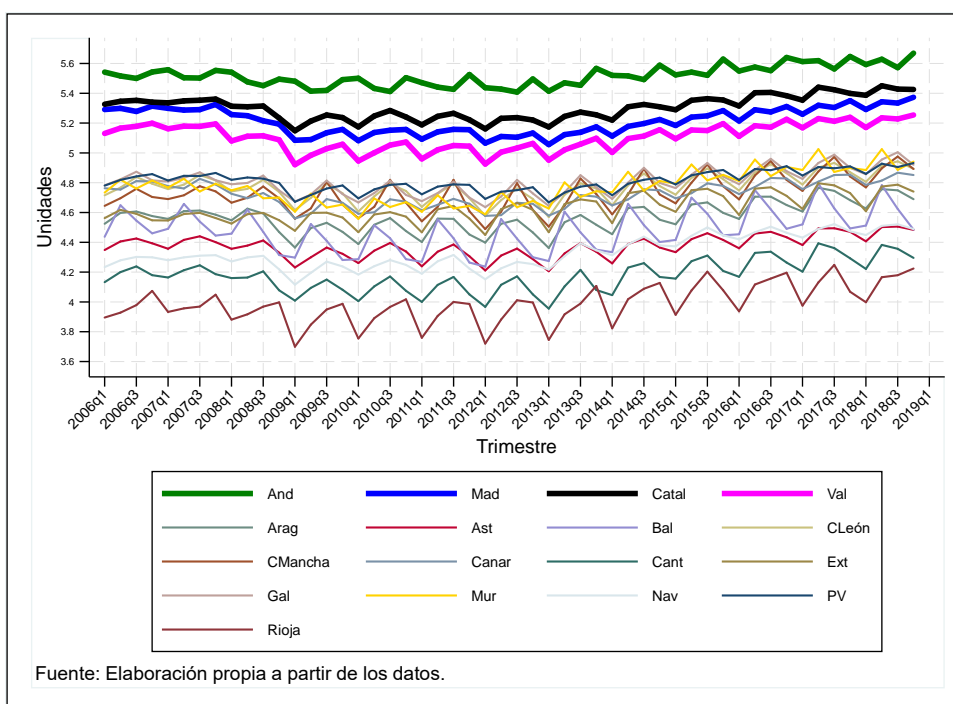
#### **4.7. Evolución trimestral del flujo de colocaciones por CCAA.**

Las colocaciones alcanzan sus máximos en la comunidad de Andalucía, seguida de Valencia, Madrid y Cataluña. Durante los años de crisis económica, el número de colocaciones desciende (variable procíclica), produciéndose un aumento a partir de 2013, año donde comienza a recuperarse la economía española. Como se ha comentado previamente, Andalucía presenta la particularidad de tener un elevado volumen relativo de contratos temporales; esto es, las series representadas en la figura 4.8 podrían cambiar significativamente si solo se representara la generación de colocaciones de calidad (empleos estables y con buenas condiciones laborales).

En comunidades como Cantabria, La Rioja y Baleares se aprecia claramente el componente estacional debido a las actividades económicas (agricultura, turismo, etc.) predominantes en estas regiones.



a) Evolución trimestral de las colocaciones por CCAA en escala lineal. Periodo 2006:T1–2018:T4.



b) Evolución trimestral de las colocaciones CCAA en escala logarítmica. Periodo 2006:T1–2018:T4.

**Figura 4.8. Evolución trimestral del flujo de colocaciones por CCAA. 2006:T1–2018:T4.**

#### 4.8. Descripción de las variables laborales en España y por CCAA. Periodo 2006:T1–2018:T4.

Previamente a la descripción de los valores de las variables laborales, es necesario introducir ciertas consideraciones aplicadas en las tasas de emparejamiento y en el grado de tensión. Para ello, se emplea la literatura de Coles y Petrongolo (2003), donde se considera el problema de agregación temporal en un modelo de emparejamiento aleatorio o “*random matching*” partiendo, en primer lugar, del supuesto siguiente:

- Los flujos de desempleados ( $uf$ ) y de vacantes ( $vf$ ) son constantes dentro de un periodo, y
- $\lambda, p$  son constantes en el periodo de tiempo.

entonces el número total de emparejamientos en el periodo considerado es

$$M_n = U_n[1 - e^{-\lambda_n}] + u_n p_n + u_n(1 - p_n) \left[ \frac{e^{-\lambda_n} - 1 + \lambda_n}{\lambda_n} \right] \quad (4.4)$$

En la expresión (4.4) se diferencian tres términos. El primero describe el stock inicial de desempleados que se emparejan exitosamente dentro del periodo, es decir, un mes; el segundo hace referencia a aquellos que han sido despedidos, pero inmediatamente encuentran trabajo; el último término muestra aquellos que han sido despedidos y más adelante, se emparejan con una nueva vacante.

Una estructura similar se aplica también para las vacantes en la ecuación (4.5). En particular, si una nueva vacante se empareja inmediatamente, dicho emparejamiento se realiza con probabilidad  $q$ . En el caso contrario, se empareja, posteriormente, a partir de una tasa  $\mu$ , quedando la siguiente expresión:

$$M_n = V_n[1 - e^{-\mu_n}] + v_n \left[ 1 - (1 - q_n) \frac{1 - e^{-\mu_n}}{\mu_n} \right] \quad (4.5)$$

La agregación temporal implica la construcción de medidas “*at risk*” para el stock de vacantes y desempleados.

Si se define  $\bar{U}_n, \bar{V}_n$  como

$$\bar{U}_n = U_n + \frac{e^{-\lambda_n} - 1 + \lambda_n}{\lambda_n [1 - e^{-\lambda_n}]} u_n, \quad (4.6)$$

$$\bar{V}_n = V_n + \frac{e^{-\mu_n} - 1 + \mu_n}{\mu_n [1 - e^{-\mu_n}]} v_n, \quad (4.7)$$



entonces  $\bar{U}_n$  puede ser interpretado como las medidas “*at risk*” de desempleados en el mes considerado  $n$ . Por ejemplo, si la tasa de emparejamiento toma un valor  $\lambda_n \approx 0$  (es decir, cada desempleado se empareja lentamente) implica  $\bar{U}_n \approx U_n + 0,5u_n$ . Debido a que ningún demandante encuentra empleo con motivo de una tasa de emparejamiento en torno a cero ( $\lambda_n \approx 0$ ), cada trabajador que se convierte en desempleado durante el mes  $n$ , permanece desempleado, de promedio, la mitad de dicho mes (suponiendo que los nuevos desempleados entran en el mercado a una tasa uniforme). Por tanto,  $U_n + 0,5u_n$  expresa el número promedio de desempleados en riesgo para el total del mes considerado. Si, por el contrario, la tasa de emparejamiento tiende a infinito, lo que implica un emparejamiento inmediato de los desempleados, entonces  $\bar{U}_n = U_n + u_n$ . Al considerar la teoría de emparejamiento aleatorio  $p_n = 0$ , la ecuación (4.4) se simplifica, resultando:

$$M_n = \bar{U}_n [1 - e^{-\lambda_n}] \quad (4.8)$$

De modo que, el emparejamiento aleatorio y la agregación temporal de los datos implica que los emparejamientos totales en un mes  $n$  se pueden expresar en el número total “*at risk*” de desempleados en ese mes, los cuales se emparejan, de promedio, a una tasa de  $\lambda_n$ . Aplicando el mismo fundamento en la ecuación (4.5) se obtiene la siguiente expresión

$$M_n = \bar{V}_n [1 - e^{-\mu_n}] \quad (4.9)$$

cuya interpretación es idéntica a la ecuación referida a los desempleados, pero en este caso, aplicada a las vacantes. Por tanto, igualando ambas expresiones se obtiene que el número de trabajadores y vacantes que se emparejan en un periodo determinado coinciden.

$$\bar{V}_n [1 - e^{-\mu_n}] = \bar{U}_n [1 - e^{-\lambda_n}] \quad (4.10)$$

En la tabla 4.1 se recogen los valores de las variables más significativas (desde el punto de vista del emparejamiento laboral) en el mercado laboral español desde el año 2006 hasta el año 2018. En referencia al número de desempleados se observa como durante el periodo estudiado la media de personas sin empleo se encuentra alrededor de los cinco millones, alcanzando un máximo de seis millones y medio, cifra correspondiente, al peor registro durante la crisis. Estas cifras son más elevadas que las correspondientes al desempleo de la EPA –la EPA cifra el desempleo actual en 3,2 millones de personas, frente a los 4,5 millones de demandantes registrados en el SEPE actualmente–, pero hay que tener en cuenta que entre

los demandantes del SEPE hay personas que la EPA no clasificaría como desempleados, al no cumplir ciertos requisitos como, por ejemplo, el de la búsqueda proactiva de empleo.

Por otro lado, los contratos iniciales se encuentran entorno a un millón y medio, dato significativo comparado con el número de desempleados analizado anteriormente. De ahí que la tasa de emparejamiento de trabajadores  $\frac{M}{U+0,5uf}$  sea de 0,28, el 28% de los demandantes consigue algún tipo de contrato cada periodo (todo ello en un contexto de mucha recurrencia en el empleo por parte de muchos trabajadores, con muchos contratos cortos o muy cortos).

En cuanto a la tasa de emparejamiento de los puestos,  $\frac{M}{V+0,5vf}$ , presenta un valor medio de 1,37. La tasa es más elevada que la de los trabajadores porque las vacantes se encuentran en el lado corto del mercado, es decir, los desempleados superan con creces a las vacantes. El que salga una tasa mayor a la unidad es un indicio de que en el flujo de colocaciones ( $M$ ) interviene un mayor porcentaje de las nuevas vacantes ( $vf$ ) que el 0,5 que hemos supuesto –más investigación sería necesaria para afinar dicho porcentaje.

Por último, el grado de tensión,  $\frac{V+0,5vf}{U+0,5uf}$ , muestra un valor cercano a 0,2 a nivel nacional, esto es debido a que el número de vacantes es muy inferior al número de desempleados. Desde el punto de vista de las empresas, la tensión es baja debido a que el “pool” de desempleados, para cubrir el escaso número de vacantes, es relativamente amplio. En cambio, entre los trabajadores surge congestión (han de competir por los escasos puestos vacantes).

**Tabla 4.1. Descripción de las variables laborales en España. 2006: T1 – 2018: T4.**

Variable	España				
	Media	Mediana	Máximo	Mínimo	Desviación Estándar
Stock de demandantes (U)	4.966.654	5.232.472	6.520.957	2.950.325	1.121.198
Contratos iniciales (C <sub>iniciales</sub> )	1.384.932	1.368.671	1.890.206	973.307	245.075
Tasa de emparejamiento de puestos	1,37	1,36	2,91	0,83	0,36
Tasa de emparejamiento de trabajadores	0,28	0,25	0,46	0,14	0,1
Grado de tensión	0,22	0,17	0,54	0,1	0,12

Fuente: Variables laborales. *Estadística de Empleo y Encuesta de Coyuntura Laboral*. SEPE y ECL.

#### 4.8.1. Descripción de las variables laborales por CCAA. 2006:01 – 2018:04.

A continuación, en la tabla 4.2 se muestra el stock de demandantes para cada una de las comunidades autónomas que componen el Reino de España. Destaca Andalucía con el mayor promedio de desempleados, tomando un valor próximo al millón y medio de demandantes, lo que supone un 25 % del total de desempleados del país, seguida de Cataluña, Valencia y Madrid, con valores en torno al medio millón de personas en desempleo. Las comunidades restantes se caracterizan por un promedio bastante inferior con respecto a las regiones anteriores, si bien son regiones con un menor volumen de población y población activa.

**Tabla 4.2. Descripción del stock de demandantes por CCAA. 2006:01 – 2018:04.**

CCAA	Stock de demandantes (U)				
	Media	Mediana	Máximo	Mínimo	Desviación Estándar
Andalucía	1.263.552	1.324.518	1.551.893	843.324	222.743
Cataluña	622.444	643.619	866.471	334.247	173.449
Valencia	551.345	592.631	746.432	266.725	151.929
Madrid	522.341	556.382	688.417	293.813	125.178
Galicia	274.642	278.980	366.067	188.851	52.767
Canarias	269.163	290.405	342.630	152.806	57.247
Castilla La Mancha	230.436	243.945	327.808	113.583	65.322
Castilla y León	225.012	231.453	311.816	137.468	50.947
País Vasco	206.631	225.237	277.888	101.190	65.500
Extremadura	175.459	177.534	223.547	126.028	29.037
Murcia	150.847	162.783	206.050	60.363	45.770
Aragón	112.349	120.776	159.196	55.181	32.831
Asturias	106.180	108.408	142.941	63.977	22.721
Baleares	98.960	98.903	155.862	38.597	32.507
Navarra	53.095	56.434	72.277	30.557	12.218
Cantabria	52.955	55.885	76.417	25.905	14.777
La Rioja	25.500	26.643	35.583	14.240	6.610

Fuente: Stock de demandantes. *Encuesta de Coyuntura Laboral*. ECL.

Si ahora analizamos los contratos iniciales por CCAA, resulta interesante el hecho de que Andalucía sea la comunidad que genera el mayor número de contratos, 333.711, y que, al mismo tiempo, sea la región que registra el número de desempleados más elevado del país. Esta situación se produce por la temporalidad característica de los contratos iniciales en esta región. Cataluña y Madrid también se encuentran a la cabeza de la generación de empleo, por detrás de Andalucía, debido la importancia de sus sectores industrial y de servicios.

**Tabla 4.3. Descripción de contratos iniciales por CCAA. 2006:01 – 2018:04.**

Contratos iniciales ( <i>M</i> )					
CCAA	Media	Mediana	Máximo	Mínimo	Desviación Estándar
Andalucía	333.711	328.512	467.253	255.876	53.512
Cataluña	204.569	204.820	282.210	140.745	36.934
Valencia	130.068	129.276	179.735	83.383	26.195
Madrid	166.646	164.021	236.282	113.695	33.381
Galicia	65.434	64.701	101.061	43.252	13.714
Canarias	53.544	53.066	73.639	36.312	10.201
Castilla La Mancha	56.001	52.592	94.600	30.818	15.686
Castilla y León	60.768	60.207	87.827	37.110	12.411
País Vasco	65.405	65.559	85.556	46.538	10.063
Extremadura	43.386	40.416	62.620	28.169	9.445
Murcia	60.669	56.877	105.971	36.084	16.391
Aragón	38.030	36.368	58.374	23.026	8.767
Asturias	24.010	23.991	32.064	16.063	4.076
Baleares	31.965	30.056	62.239	17.269	10.775
Navarra	22.274	20.329	33.269	13.068	5.700
Cantabria	15.426	14.780	24.729	9.022	3.870
La Rioja	10.175	9.705	17.722	5.008	3.033

Fuente: Contratos iniciales. *Estadística de Empleo*. SEPE.

Por su parte, en la tabla 4.4 se observan las tasas de emparejamientos de puestos vacantes a nivel autonómico, cuya fórmula viene dada por:  $C_{inic}/(V + 0,5vf)$ . A excepción de Cataluña y Galicia, el valor promedio de las tasas de emparejamiento de vacantes supera la unidad (esto es,  $C_{inic} > (V + 0,5vf)$ ). Esta desigualdad es extraña si tenemos en cuenta que el número máximo de colocaciones en cada periodo viene limitado por el mínimo entre dos cifras: la del desempleo y la de vacantes –por ejemplo, si en la economía hay 100 desempleados y 50 vacantes, el número máximo de colocaciones que se pueden generar en el periodo son 50. Como en España el número de vacantes es muy inferior al de desempleados, debería cumplirse que  $C_{inic} \leq (V + 0,5vf)$ . Por tanto, la desigualdad  $C_{inic} > (V + 0,5vf)$  implica necesariamente que, en el emparejamiento de cada trimestre, no está participando solo el 50% del flujo de nuevas vacantes, sino un porcentaje bastante superior –tal y como se ha comentado previamente, una investigación exhaustiva sería necesaria para determinar un porcentaje más preciso de nuevas vacantes. Entre las comunidades con una tasa de emparejamiento de puestos más elevada se encuentran Andalucía, Castilla La Mancha y Extremadura, entre otras. En el caso de Andalucía, a pesar de ser la comunidad con mayor número de desempleados, presenta una alta tasa de emparejamiento debido, como se comentó anteriormente, a la elevada temporalidad de los contratos generados. Téngase en

cuenta que en las tres CCAA comentadas, la agricultura tiene un peso importante, siendo este sector uno de los que mayor temporalidad presenta.

**Tabla 4.4. Descripción de la tasa de emparejamiento de puestos por CCAA. 2006:01 – 2018:04.**

$Tasa\ de\ emparejamiento\ de\ puestos = \frac{C_{mic}}{V_{total} + 0,5 \cdot v_{f_{total}}}$					
CCAA	Media	Mediana	Máximo	Mínimo	Desviación Estándar
Andalucía	1,54	1,59	2,85	0,51	0,43
Cataluña	0,71	0,55	4,34	0,21	0,69
Valencia	1,44	1,25	5,11	0,41	0,85
Madrid	1,33	1,23	2,92	0,64	0,45
Galicia	0,92	0,86	2,78	0,24	0,43
Canarias	1,2	1,09	3,32	0,31	0,57
Castilla La Mancha	1,58	1,47	3,51	0,36	0,65
Castilla y León	1,23	1,18	4,21	0,45	0,57
País Vasco	1,35	1,15	5,75	0,55	0,89
Extremadura	1,58	1,69	2,37	0,65	0,41
Murcia	1,44	1,31	4,74	0,72	0,72
Aragón	1,51	1,5	4,01	0,73	0,54
Asturias	1,53	1,46	6	0,44	0,84
Baleares	1,43	1,43	2,38	0,47	0,42
Navarra	1,13	1,05	4,12	0,53	0,48
Cantabria	1,45	1,18	8,86	0,47	1,23
La Rioja	1,42	1,34	4,29	0,81	0,54

Fuente: Tasa de emparejamiento de puestos. *Estadística de Empleo y Encuesta de Coyuntura Laboral*. SEPE y ECL.

A diferencia de las tasas de emparejamiento de puestos, las tasas de emparejamiento de trabajadores mostradas en la tabla 4.5 –que relacionan las colocaciones ( $M$ ) con el stock de desempleados y los nuevos demandantes que surgen durante el periodo analizado:  $C_{mic}/(U + 0,5uf)$  –toman valores inferiores, por debajo de la unidad, en todas las comunidades autónomas. Esta situación implica que los trabajadores en desempleo superan a los puestos vacantes, por lo que las vacantes se encuentran en el lado corto del mercado, mostrando por tanto mejores tasas de emparejamiento. Los valores más altos de tasa de emparejamiento de trabajadores se observan en Murcia, Navarra y La Rioja. Haría falta más investigación para determinar las causas de la posición favorable de estas regiones. En cualquier caso, se trata de tres regiones medianas o pequeñas, dos de ellas con un nivel de renta relativamente elevado (Navarra y la Rioja), y la otra (Murcia) con el crecimiento del PIB relativamente fuerte en el periodo analizado.

**Tabla 4.5. Descripción de la tasa de emparejamiento de trabajadores por CCAA. 2006:01 – 2018:04.**

$Tasa\ de\ emparejamiento\ de\ trabajadores = \frac{C_{Inic}}{U_s + 0,5 \cdot u_f}$					
CCAA	Media	Mediana	Máximo	Mínimo	Desviación Estándar
Andalucía	0,26	0,24	0,38	0,16	0,07
Cataluña	0,34	0,3	0,61	0,16	0,15
Valencia	0,25	0,21	0,54	0,11	0,12
Madrid	0,33	0,28	0,63	0,16	0,15
Galicia	0,24	0,22	0,43	0,11	0,08
Canarias	0,2	0,17	0,39	0,11	0,08
Castilla La Mancha	0,25	0,24	0,47	0,09	0,1
Castilla y León	0,27	0,25	0,47	0,11	0,09
País Vasco	0,33	0,27	0,65	0,18	0,14
Extremadura	0,23	0,22	0,36	0,12	0,06
Murcia	0,43	0,37	0,99	0,19	0,22
Aragón	0,35	0,29	0,66	0,14	0,16
Asturias	0,23	0,21	0,39	0,11	0,07
Baleares	0,36	0,29	0,91	0,11	0,24
Navarra	0,41	0,41	0,65	0,2	0,14
Cantabria	0,3	0,26	0,62	0,11	0,13
La Rioja	0,4	0,37	0,73	0,15	0,16

Fuente: Tasa de emparejamiento de trabajadores. *Estadística de Empleo*. SEPE.

Otra variable de interés en el análisis del emparejamiento laboral es el grado de tensión, que relaciona las vacantes generadas en una CCAA y el total de desempleados registrados en la misma comunidad. Observando los valores mostrados en la tabla 4.6, la región más destacable es Cataluña, con un valor medio durante el periodo analizado de 0,69. El resto de las comunidades presentan un grado de tensión o ratio vacantes/desempleo inferior, lo que implica un menor número de vacantes por desempleado. Por tanto, desde la posición del trabajador, existe una congestión, ya que se produce una mayor competencia entre los desempleados por conseguir las vacantes ofrecidas la región. Desde la perspectiva de la demanda de trabajo, las empresas se encuentran en un amplio abanico de trabajadores donde elegir, por tanto, les resulta más sencillo e inmediato encontrar un candidato acorde a la vacante a cubrir.

**Tabla 4.6. Descripción del grado de tensión por CCAA. 2006:01 – 2018:04.**

$\text{Grado de tensión} = \frac{V_{total} + 0,5 \cdot v_{f_{total}}}{U_s + 0,5 \cdot u_f}$					
CCAA	Media	Mediana	Máximo	Mínimo	Desviación Estándar
Andalucía	0,19	0,16	0,64	0,08	0,1
Cataluña	0,69	0,52	2,73	0,13	0,57
Valencia	0,22	0,17	0,71	0,06	0,16
Madrid	0,27	0,22	0,77	0,12	0,16
Galicia	0,32	0,23	1,27	0,09	0,23
Canarias	0,21	0,16	1,27	0,07	0,18
Castilla La Mancha	0,18	0,14	0,48	0,06	0,11
Castilla y León	0,25	0,22	0,83	0,07	0,13
País Vasco	0,27	0,25	0,62	0,08	0,1
Extremadura	0,16	0,14	0,4	0,07	0,07
Murcia	0,33	0,27	1,32	0,12	0,2
Aragón	0,27	0,19	0,9	0,09	0,18
Asturias	0,17	0,16	0,65	0,04	0,09
Baleares	0,26	0,22	0,98	0,08	0,18
Navarra	0,39	0,35	0,65	0,11	0,15
Cantabria	0,27	0,22	0,93	0,04	0,17
La Rioja	0,29	0,26	0,57	0,1	0,11

Fuente: Grado de tensión. *Estadística de Empleo y Encuesta de Coyuntura Laboral*. SEPE y ECL.

#### 4.9. Descripción de otras variables relevantes del mercado del trabajo en España. 2006:01 – 2018:04.

Resulta interesante analizar a que sectores están vinculados los contratos realizados en España, lo cual puede ayudar a entender los resultados observados en el apartado anterior. En la tabla 4.7, se muestra que, en comunidades como Andalucía, Extremadura, Murcia y Castilla La Mancha, predominan los contratos en actividades pertenecientes sector primario, concretamente, la agricultura.

Madrid y Cataluña, a pesar de poseer un sector industrial referente en España, no generan un porcentaje de contratos significativos en este sector si lo comparamos con el sector servicios, ya que, en ambas regiones destacan, al igual que en el resto de las regiones, las actividades relacionadas con las finanzas, la hostelería o el turismo, entre otras –la economía española está terciarizada y presenta un sector cuaternario cada vez más importante.

**Tabla 4.7. Estructura de la generación de empleo. Valores medios trimestrales. 2009: T1-2018: T4.**

CCAA	Porc. contratos Agricultura	Porc. contratos Industria	Porc. contratos Construcción	Porc. contratos Servicios
Andalucía	32,0%	4,7%	9,9%	53,4%
Aragón	11,5%	13,1%	5,9%	69,5%
Asturias	1,0%	9,5%	6,3%	83,1%
Islas Baleares	1,1%	3,0%	11,8%	84,1%
Castilla y León	6,2%	13,5%	6,2%	74,0%
Castilla La Mancha	21,3%	9,2%	8,7%	60,8%
Canarias	2,0%	3,3%	7,9%	86,9%
Cantabria	0,9%	13,4%	6,2%	79,5%
Cataluña	2,7%	9,4%	5,7%	82,2%
Extremadura	42,5%	4,0%	9,9%	43,7%
Galicia	3,1%	11,6%	6,4%	78,9%
Madrid	0,5%	4,7%	7,9%	87,0%
Murcia	31,0%	6,9%	5,6%	56,5%
Navarra	3,9%	19,8%	4,4%	71,8%
País Vasco	1,9%	11,9%	4,9%	81,2%
La Rioja	21,8%	15,2%	5,1%	57,9%
Valencia	8,6%	10,4%	6,9%	74,1%

Fuente: Contratos iniciales. *Estadística de Empleo*. SEPE.

Este apartado concluye con los valores de la actividad laboral y de la temporalidad en el empleo mostrados en la tabla 4.8, valores procedentes de la *Encuesta de Población* (EPA). Se puede destacar la temporalidad de los contratos en Andalucía, Extremadura y Murcia, siendo las comunidades con mayores porcentajes de contratación temporal –muestran porcentajes de ocupados temporales superiores al 33%. Por el contrario, las comunidades con el máximo número de ocupados con contrato fijo son Madrid y Cataluña, debido posiblemente a sus desarrollados sectores industriales; además de ser comunidades con un sector turismo relativamente estable, permitiendo dicha estabilidad que los servicios turísticos y hosteleros generen un cierto número de contratos fijos. A su vez, las actividades financieras suponen un papel importante dentro del sector servicios en las regiones de Madrid y Cataluña.



**Tabla 4.8. Estructura del mercado de trabajo y temporalidad en el stock de ocupados.  
2009: T1-2018: T4.**

CCAA	Población total	Activos	Ocupados	Ocupados Fijos	Ocupados Temporales	Porc. Ocupados Fijos	Porc. Ocupados Temporales
Andalucía	6852,1	3994,2	2301,8	1505,9	795,9	65,4%	<b>34,6%</b>
Cataluña	6140,3	3855,9	2649,9	2129,2	520,7	80,3%	19,7%
Madrid	5293,1	3412,4	2518,1	2060,4	457,8	81,8%	18,2%
Valencia	4132,5	2459,3	1560,7	1130,2	430,5	72,4%	27,6%
Galicia	2374,7	1283,1	828,8	621,0	207,8	74,9%	25,1%
Castilla y León	2119,2	1164,7	765,6	591,5	174,1	77,3%	22,7%
País Vasco	1829,1	1049,9	768,4	591,3	177,0	77,0%	23,0%
Canarias	1771,4	1092,5	673,0	468,2	204,8	69,6%	30,4%
Castilla La Mancha	1700,2	1000,0	613,1	451,0	162,1	73,6%	26,4%
Murcia	1182,9	717,3	459,8	307,7	152,1	66,9%	33,1%
Aragón	1109,1	655,4	452,5	346,7	105,8	76,6%	23,4%
I. Baleares	930,1	609,3	403,9	296,4	107,5	73,4%	26,6%
Asturias	923,4	477,0	315,6	238,6	77,0	75,6%	24,4%
Extremadura	912,4	500,9	291,4	190,7	100,8	65,4%	<b>34,6%</b>
Navarra	523,9	313,8	225,5	172,2	53,4	76,3%	23,7%
Cantabria	496,9	279,2	193,5	150,2	43,3	77,6%	22,4%
La Rioja	262,9	157,0	108,0	84,6	23,3	78,4%	21,6%

Fuente: *Encuesta de Población Activa*. INE.

## 5. ESTIMACIÓN DEL MODELO ECONÓMICO Y PRINCIPALES RESULTADOS.

En la presente sección se estima la eficiencia de la función de emparejamiento en el mercado laboral español, a partir de la metodología de datos de panel, empleando las series proporcionadas fundamentalmente por el SEPE en el periodo temporal 2009-2018 –los datos anteriores a 2009 presentan problemas de representatividad debido a los cambios metodológicos en el sistema de recogida de información en el SEPE. Para ello, se aplican los estimadores de datos de panel descritos en secciones anteriores, concretamente, el estimador de efectos fijos y efectos aleatorios. La elección de uno u otro como el estimador más adecuado (consistente) se fundamenta en el test de Hausman, el cual se describe en el siguiente apartado.

La función de emparejamiento a estimar, objeto de este Trabajo Fin de Máster, mantiene la siguiente forma:

$$\ln M_{it} = \alpha_0 + \alpha_1 \ln U_{it-1} + \alpha_2 \ln uf_{it} + \alpha_3 \ln V_{it-1} + \alpha_4 \ln vf_{it} + \alpha_5 P + \alpha_6 D_{ag} + \alpha_7 D_{serv} + \alpha_8 D_{const} + \alpha_9 T_1 + \alpha_{10} T_1 + \alpha_{11} T_2 + \alpha_{12} T_3 + u_i + \varepsilon_{it} \quad (5.1)$$

Dentro de la expresión (5.1) se diferencian dos subíndices,  $i$ , que expresa la unidad muestral o el individuo –en nuestro caso, hace referencia a la comunidad autónoma– y  $t$ , que indica el trimestre del año en el periodo de tiempo considerado (2009-2018). La variable a explicar viene dada por el flujo de colocaciones producido a lo largo de cada trimestre. Las variables explicativas introducidas en el modelo se corresponden con las series de datos al inicio de cada trimestre (o al final del trimestre anterior) de los stocks de vacantes ( $V$ ) y demandantes ( $U$ ), con los flujos producidos durante el trimestre de nuevos desempleados ( $uf$ ) y nuevas vacantes ( $vf$ ). Obsérvese que los stocks de vacantes y de desempleados al final del trimestre,  $V_{it}$  y  $U_{it}$ , entran en el modelo retrasados un trimestre ( $V_{it-1}$  y  $U_{it-1}$ ), de esta forma ambas variables están midiendo los respectivos stocks al final del trimestre  $t - 1$ , o lo que es lo mismo, al comienzo del trimestre  $t$ .

Aparte de las variables sobre desempleados y vacantes, la estimación contiene una serie de variables dummy para controlar aspectos del mercado de trabajo que pudieran influir en la eficiencia regional del proceso de emparejamiento. Así, se ha incluido una dummy que controla el porcentaje de empleo temporal en cada CCAA, tres dummies que permiten medir el efecto de la estructura porcentual sectorial del empleo en cada CCAA  $\{D_{ag}, D_{serv}, D_{const}\}$  (la dummy correspondiente a la industria se omite del modelo estimado para evitar posibles

problemas de multicolinealidad), y tres dummies que permiten controlar el trimestre (dentro de cada año) en que nos encontramos. Finalmente,  $\varepsilon_{it}$  es el término de error correspondiente a cada observación y  $u_i$  representa el efecto individual correspondiente a cada CCAA y que nos permite controlar su heterogeneidad inobservable. Obsérvese que como el modelo es logarítmico-logarítmico en las variables  $\{U_{it}, u_{fit}, V_{it}, v_{fit}\}$ , los respectivos coeficientes estimados  $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$  representan elasticidades matemáticas.

### **5.1. Comparativa entre el estimador de efectos fijos y el de efectos aleatorios (Test de Hausman).**

Los modelos de efectos fijos y de efectos aleatorios generan los estimadores más habituales en la estimación de los datos de panel, es por ello que, mediante el test de Hausman se realiza un proceso comparativo de ambos estimadores, para determinar qué modelo presenta una mayor consistencia y eficiencia, de cara a obtener una estimación de la función de emparejamiento más próxima a la realidad poblacional.

El test de Hausman es un contraste basado en una distribución Chi cuadrado  $X^2$  que permite detectar si existe correlación entre el efecto individual  $u_i$  y las variables explicativas del modelo. La hipótesis nula ( $H_0$ ) del contraste de Hausman es la igualdad, que determina si los coeficientes del estimador de efectos fijos y de efectos aleatorios presentan una diferencia sistemática entre ellos o no. Si los coeficientes difieren sistemáticamente (tienen un sesgo definido), se puede concluir que existe correlación entre los regresores del modelo y el término del efecto individual, entonces se procede a rechazar la hipótesis nula y se emplea, por defecto, el modelo de efectos fijos, ya que ofrece un estimador consistente bajo la existencia de dicha correlación –el modelo de efectos fijos es siempre consistente, bajo la hipótesis nula y bajo la hipótesis alternativa; el modelo de efectos aleatorios resulta inconsistente bajo la hipótesis alternativa, pero es consistente y más eficiente que el modelo de efectos fijos bajo la hipótesis nula.

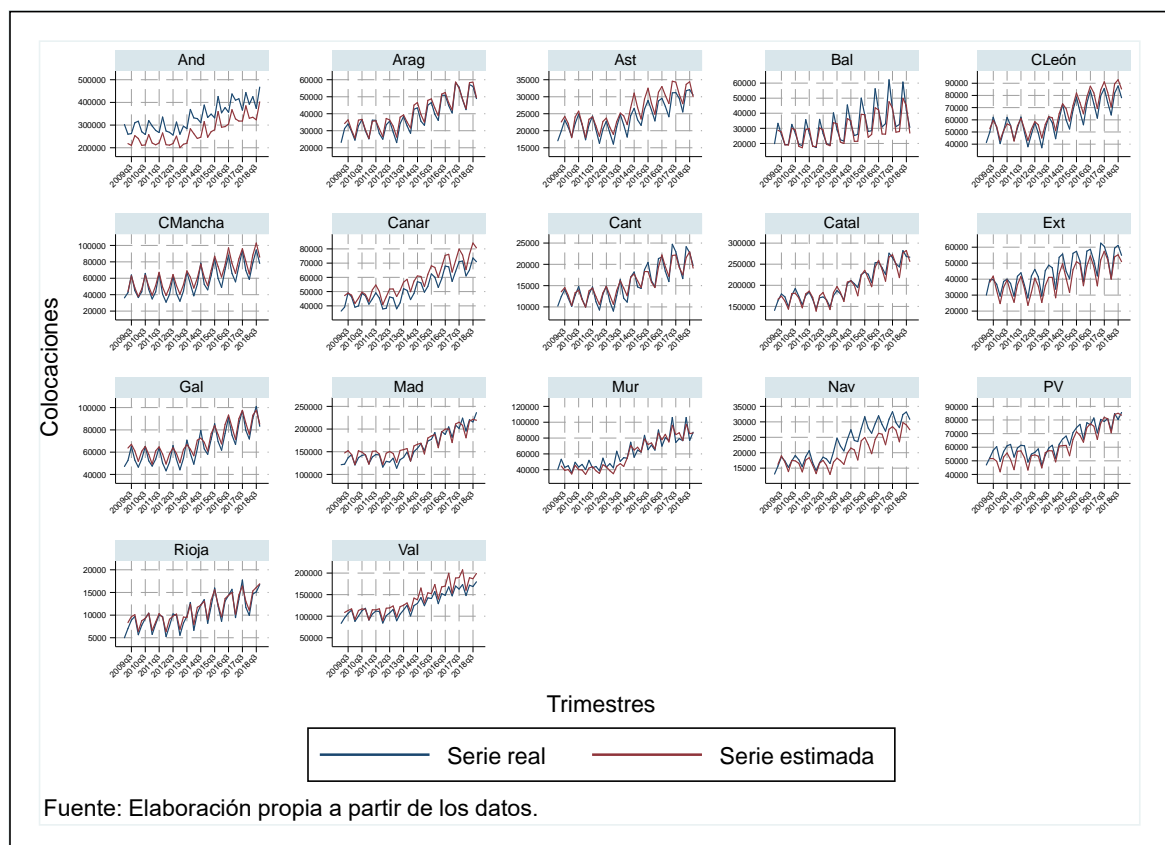
Al ejecutar el test de Hausman al modelo estimado (5.1) (con las series de datos correspondientes al periodo 2009-2018) se obtiene un valor negativo del estadístico chi cuadrado (-27,93) que carece de sentido dada la distribución positiva que sigue este estadístico. La incongruencia de un valor negativo del estadístico  $X^2$  puede ser debida a un tamaño relativamente pequeño de la muestra (Baum, 2006). En estos casos, lo habitual es estimar el panel de datos por efectos fijos para controlar la posible existencia de correlación entre los regresores y el término de efectos individuales.

## 5.2. Estimación agrupada de la función de emparejamiento a partir de la metodología de datos de panel.

Con la finalidad de mostrar el efecto de emplear el modelo de efectos fijos en la metodología de datos de panel para evaluar la función de emparejamiento definida al inicio de la sección 5, es conveniente realizar una estimación previa, empleando el modelo más genérico y simple de datos de panel consistente en ignorar el término de efecto individual de cada CCAA; es decir, omitimos del modelo la variable  $u_i$ , que captura la heterogeneidad inobservable que pueda existir en cada región.

$$\ln M_{it} = \alpha_0 + \alpha_1 \ln U_{it-1} + \alpha_2 \ln uf_{it} + \alpha_3 \ln V_{it-1} + \alpha_4 \ln vf_{it} + \alpha_5 P + \alpha_6 D_{ag} + \alpha_7 D_{serv} + \alpha_8 D_{const} + \alpha_9 T_1 + \alpha_{10} T_2 + \alpha_{11} T_3 + \varepsilon_{it} \quad (5.2)$$

Una vez que estimamos el panel (5.2), podemos comparar las colocaciones reales de cada comunidad durante el periodo 2009-2018 con la predicción de dicha variable entregada por el modelo estimado, dicha comparación se representa en la figura 5.1.



**Figura 5.1. Estimación de la función de emparejamiento a partir del modelo general de datos de panel. Periodo 2009:T1-2018:T4.**

Por un lado, se observan las series de colocaciones reales (línea azul) desagregadas por CCAA, series extraídas de la *Estadística de empleo* del Servicio Público de Empleo Estatal. Por otro lado, la línea roja representa las series de colocaciones estimadas mediante el modelo agrupado (*pool regression*) de datos de panel. El hecho de que el modelo falle bastante en predecir el flujo de contrataciones en algunas CCAA (como, por ejemplo, en Andalucía o Navarra) se debe a que no está teniendo en cuenta la heterogeneidad no observable de cada una de ellas; recuérdese que el modelo agrupado, asume que todas las regiones son iguales (presentan condiciones similares en el mercado de trabajo) y, por tanto, no existen diferencias entre sí.

La figura 5.1 muestra que, en Andalucía, el modelo general predice un flujo de colocaciones, dados los valores de las variables explicativas, inferior al flujo real de colocaciones existente. En otras palabras, Andalucía genera bastante más empleo de lo que el propio modelo predice. A esta diferencia entre la serie real y la estimada se le denomina residuo.

### **5.3. Estimación de la función de emparejamiento aplicando el estimador de efectos fijos.**

Al aplicar el estimador de efectos fijos sobre la función de emparejamiento –modelo (5.1)–, se obtienen las estimaciones recogidas en la tabla 5.1, para un total de 17 comunidades autónomas o unidades muestrales y 663 observaciones, es decir, una observación por trimestre y CCAA a lo largo del periodo de tiempo objeto de estudio: 2009:T2-2018:T4 –el primer trimestre de 2009 se pierde en la estimación porque los stocks de vacantes y desempleados entran en el modelo con un retardo de un periodo . Las filas de la tabla se corresponden con las variables explicativas introducidas en la función de emparejamiento; en las columnas se contemplan los resultados obtenidos para el periodo analizado completo y para dos subperiodos bien diferenciados: 2009-2012, intervalo correspondiente al período de crisis económica de la economía española, y 2013-2018, subperiodo de recuperación económica.

**Tabla 5.1. Resultados de la función de emparejamiento estimada por el modelo de efectos fijos. Periodo 2009:T1-2018:T4.**

Variables explicativas	Información estimada	MODELO EFECTOS FIJOS		
		2009-2018	2009-2012 Disminuyen Contr.	2013-2018 Aumentan Contr.
Stock de demandantes (-1, log)	Coefficiente estimado	0.19***	0.23***	0,05
	Estadístico t ( $H_0: \beta = 0$ )	5,77	3,35	1,02
	p-valor	0	0	0,31
Nuevos demandantes (log)	Coefficiente estimado	-0.16***	-0.15***	-0.10***
	Estadístico t	7,33	-3,41	-2,95
	p-valor	0	0	0
Stock de vacantes (-1, log)	Coefficiente estimado	0.03***	0.03***	0.04***
	Estadístico t	5,31	4,09	3,92
	p-valor	0	0	0
Nuevas vacantes (log)	Coefficiente estimado	0.03***	0.02**	0.07***
	Estadístico t	3,58	2,03	5,09
	p-valor	0	0,04	0
Porcentaje contratos temporales	Coefficiente estimado	0.96***	1.06***	2.27***
	Estadístico t	4,51	3,17	6,65
	p-valor	0	0	0
Porcentaje contratos en la agricultura	Coefficiente estimado	0.48***	3.62***	0,11
	Estadístico t	4,51	6,06	0,7
	p-valor	0	0	0,48
Porcentaje contratos en los servicio	Coefficiente estimado	-0.51***	2.38***	-0.99***
	Estadístico t	4,94	3,69	-6,32
	p-valor	0	0	0
Porcentaje contratos en la construcción	Coefficiente estimado	-2.21***	1.41**	-1.96***
	Estadístico t	9,81	2,04	-4,83
	p-valor	0	0,04	0
Dummy de estacionalidad 1er trimestre	Coefficiente estimado	-0.13***	-0.11***	-0.12***
	Estadístico t	13,31	-7,05	-7,85
	p-valor	0	0	0
Dummy de estacionalidad 2º trimestre	Coefficiente estimado	0.02**	0,01	0.04***
	Estadístico t	2,42	0,42	3,18
	p-valor	0,02	0,68	0
Dummy de estacionalidad 3er trimestre	Coefficiente estimado	0.05***	0.07***	0.03***
	Estadístico t	6,63	5,76	2,65
	p-valor	0	0	0,01
Tendencia	Coefficiente estimado	-0.10***		
	Estadístico t	10,45		
	p-valor	0		
Tendencia <sup>2</sup>	Coefficiente estimado	0.00***		
	Estadístico t	11,38		
	p-valor	0		
Tendencia <sup>3</sup>	Coefficiente estimado	-0.00***		
	Estadístico t	10,9		
	p-valor	0		
Constante	Coefficiente estimado	6.20***	1,33	1,49
	Estadístico t	7,29	0,65	1,06
	p-valor	0	0,52	0,29
Nº observaciones		663	255	408
Unidades muestrales		17	17	17
R <sup>2</sup> intra-grupos		0,92	0,86	0,86
R <sup>2</sup> entre-grupos		0,99	0,99	0,98
R <sup>2</sup> global		0,98	0,98	0,97
Estadístico F ( $H_0$ : todos los coeficientes = 0)		478.31 (p 0.00)	111.36 (p 0.00)	192.63 (p 0.00)
Estadístico F ( $H_0$ : $u_i = 0$ )		46.37 (p 0.00)	20.43 (p 0.00)	23.01 (p 0.00)
corr ( $u_i, bX$ )		0,85	0,26	-0,95
$\sigma_u$		0,22	0,11	0,45
$\sigma_e$		0,07	0,06	0,08
$\rho$ (fracción de la varianza debida a $u_i$ )		0,91	0,78	0,97

Fuente: Elaboración propia.

Destacar que los modelos econométricos miden la relación entre las variables en condiciones *ceteris paribus*, de modo que, al discutir el impacto de una variable explicativa sobre la variable endógena, el resto de las variables explicativas permanecen constante. Asimismo, el ajuste del modelo, medido con el coeficiente de determinación  $R^2$ , es más elevado cuando la especificación es logarítmica-logarítmica, lo cual implica que la relación entre las colocaciones y las variables de vacantes y desempleados es no lineal. En efecto, si ambas variables, regresor y regresando, se expresan como logaritmos neperianos, el coeficiente de la variable explicativa representa una elasticidad, esto es, mide la variación porcentual de la variable independiente, en nuestro caso, el flujo de colocaciones ( $M$ ), al aumentar o disminuir porcentualmente la variable explicativa.

La primera variable explicativa estimada es el stock de demandantes ( $U$ ), caracterizada por un coeficiente positivo y significativo al 99% de confianza, de 0,19 para el periodo completo, valor que significa que si el número de demandantes se duplica en una determinada comunidad autónoma (incremento del 100%), el flujo de colocaciones solo aumentaría un 19%. Obsérvese que, al tratarse de un término de elasticidad, se estima el ratio de variaciones porcentuales  $\frac{\Delta(\%) M_t}{\Delta(\%) U_{t-1}} = 0,19$ , o expresado de otra forma,  $crec(\%)M_t = 0,19 crec(\%)U_{t-1}$ . En cualquier caso, el signo positivo del coeficiente indica que aquellas comunidades con más desempleados al comienzo de un trimestre generan más contratos iniciales durante dicho trimestre; el razonamiento se puede hacer en términos entre-grupos: aquellas regiones donde la media de trabajadores en búsqueda de empleo es mayor, generan mayores valores promedio de contratos.

Durante el subperiodo de expansión el coeficiente del stock de demandantes toma un valor de 0,05 y resulta no significativo, por lo que podemos deducir que el stock de demandantes pierde influencia en la generación de empleo en dicho subperiodo. Todo parece indicar que los demandantes que se han ido agregando al desempleo durante el periodo de crisis tienen que competir posteriormente por las nuevas vacantes que se van generando al ritmo de la recuperación económica, por lo que son estas últimas realmente las que guían la generación de empleo en el periodo expansivo.

En cuanto al flujo de nuevos demandantes ( $uf$ ), los coeficientes estimados toman valores negativos y significativos al 99% en ambos subperiodos, tanto en la fase de recesión como en la etapa de expansión económica. De la misma forma que para el stock de desempleados, el coeficiente de flujo de nuevos desempleados representa una elasticidad,  $\frac{\Delta(\%) M_t}{\Delta(\%) uf_t}$ , de manera que, si el flujo de desempleados se duplica, el flujo de colocaciones caerá

un 15% y 10% en los subperiodos 2009-2012 y 2013-2018, respectivamente. Estos valores negativos en los coeficientes estimados pueden ser un indicio de la existencia de congestión en el emparejamiento por parte de los nuevos demandantes, que tienen que competir por las relativamente escasas vacantes no solo entre ellos, sino también con el stock de demandantes; además, aquellos nuevos desempleados que no logren emparejarse durante el trimestre en que comienzan su búsqueda pasarán a formar parte del stock de desempleados.

Los coeficientes de las variables correspondientes al stock de vacantes ( $V$ ) y al flujo de nuevas vacantes ( $vf$ ) muestran coeficientes positivos, pero relativamente pequeños, generando por tanto una relación inelástica con el flujo de colocaciones ( $M$ ). El escaso efecto de las vacantes se puede deber a que existe un grado importante de desajuste o “*mismatch*” en el mercado de trabajo español, originado por la falta de coincidencia entre los requerimientos de las vacantes establecidos por las empresas y las cualificaciones aportadas por los trabajadores.

En el caso del stock de vacantes, el coeficiente obtenido durante la crisis económica resultar ser de 0,03, mientras que para el periodo comprendido entre los años 2013 y 2018, toma un valor de 0,04; ambos coeficientes positivos y significativos con un nivel de confianza del 99%. Estos coeficientes indicarían que, duplicando el stock inicial de vacantes, se consigue incrementar el flujo de colocaciones entre un 3% y 4%, dependiendo del periodo.

Los coeficientes de las nuevas vacantes resultan parecidos a los del stock de vacantes de modo que un incremento del 100% en  $vf$  dará lugar a un incremento de las colocaciones de entre el 2% y el 7%, dependiendo del subperiodo considerado.

A diferencia de las variables anteriores, que mantenían una relación logarítmica con la variable independiente, de la forma,  $\frac{\partial \ln(Y)}{\partial \ln(X)}$ , las variables que se analizan a continuación destacan por una relación logarítmica-porcentaje,  $\frac{\partial \ln(Y)}{\partial X(\%)}$ . Trasladado a nuestro modelo, la variable independiente se corresponde con el flujo de colocaciones ( $M$ ), mientras que las variables explicativas incluyen las variables de porcentajes de contratos temporales, de contratos en la agricultura, de contratos en el sector servicio y en el sector de la construcción –la variable que mide el porcentaje de las colocaciones que se producen en la industria (en cada CCAA) se ha omitido del modelo estimado para evitar problemas de multicolinealidad.

Para el conjunto del periodo, y para el subperiodo de crisis, la estimación del coeficiente de la variable porcentajes de contratos temporales toma un valor positivo significativo a un nivel de confianza del 99% y cercano a 1. Esta relación positiva implica que, si de una comunidad autónoma a otra (entre-grupos) o de un trimestre a otro (intra-



grupos), el porcentaje de contratos temporales aumenta en un punto básico o porcentual, las colocaciones crecen aproximadamente en el mismo porcentaje, resultado esperable si se tiene en cuenta la importancia que tiene la contratación temporal en la economía española. Si se observa la época de expansión económica, comprendida entre los años 2013-2018, el coeficiente toma un valor aún mayor (2,27), la mayor generación de empleo en épocas de expansión puede estar detrás de este mayor valor del coeficiente.

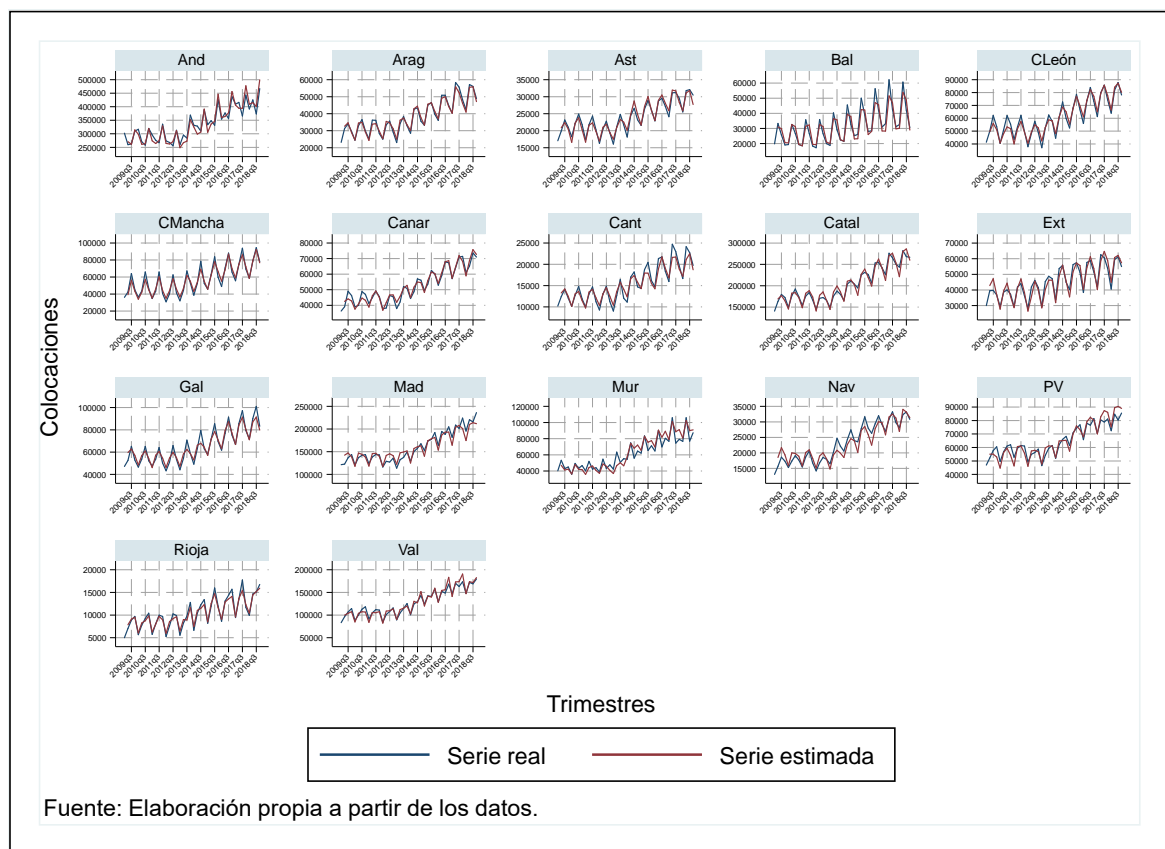
En lo que respecta a las variables porcentaje de contratos en la agricultura y en el sector servicios se produce una situación temporal diferente a la descrita anteriormente para la contratación temporal. Durante la crisis económica, una parte importante de los desempleados recurrieron a los sectores agrícola y de servicios, de manera que aquellas regiones con mayor peso de estos sectores pudieron contribuir en mayor medida a la generación de empleo en el periodo de crisis; esto explica los elevados coeficientes estimados por el modelo para estas variables en dicho periodo recesivo (3,62\*\*\* para la agricultura y 2,38\*\*\* para el sector servicios). A lo largo de la expansión económica (2013-2018), es el peso del sector industrial en cada Comunidad la variable determinante en la mayor generación relativa de empleo –obsérvese que los coeficientes del resto de sectores se han vuelto negativos o no significativos en dicho periodo expansivo.

De las variables dummies estacionales cabe destacar que permiten controlar la estacionalidad de la serie en los diferentes cuatrimestres del periodo considerado. Por otro lado, el término de tendencia permite capturar el comportamiento a largo plazo de la serie objeto de estudio; el resto de las variables explicativas  $\{U_{it}, uf_{it}, V_{it}, vf_{it}\}$  explicarían la variabilidad intra anual de la serie de colocaciones, es decir, el comportamiento a corto plazo.

Una vez descritas las variables explicativas de la función de emparejamiento, nos podemos detener en los estadísticos de contraste que ofrece el panel estimado. Entre ellos, se observan dos test correspondientes al estadístico  $F$ ; el primer test consiste en un contraste de significatividad conjunta de todas las variables explicativas, donde se plantea la hipótesis nula ( $H_0$ : todos los coeficientes toman un valor igual a cero); adicionalmente, la segunda prueba declara en la hipótesis nula la no existencia de efectos individuales, es decir, se somete a contraste que  $u_i = 0$  para todo  $i$ . Debido a los valores significativos del estadístico  $F$ , en ambos casos, se rechazan las hipótesis nulas planteadas; por tanto, se concluye que los coeficientes de las diferentes variables independientes resultan distintos de cero y, además, que cada CCAA presenta singularidades propias o efectos individuales –el panel de datos de efectos fijos asume que dichas singularidades regionales permanecen estables en el tiempo; modelos más avanzados de datos de panel rompen este supuesto.

Por su parte, la fracción de la varianza debida al término de efectos individuales  $u_i$ , definida como  $\rho$ , es otro parámetro relevante. Los resultados obtenidos del factor  $\rho$  atribuyen, en más de un 78%, las diferencias entre las regiones a la idiosincrasia de cada una de ellas y no a la existencia de perturbaciones aleatorias dentro del modelo.

Discutidos e interpretados los resultados numéricos de las variables explicativas al aplicar el estimador de efectos fijos, resulta interesante analizar gráficamente algunos aspectos del modelo estimado. Así, en la figura 5.2 se muestran las series reales de colocaciones o contratos iniciales por comunidades autónomas frente a las series estimadas a partir del modelo de efectos fijos.

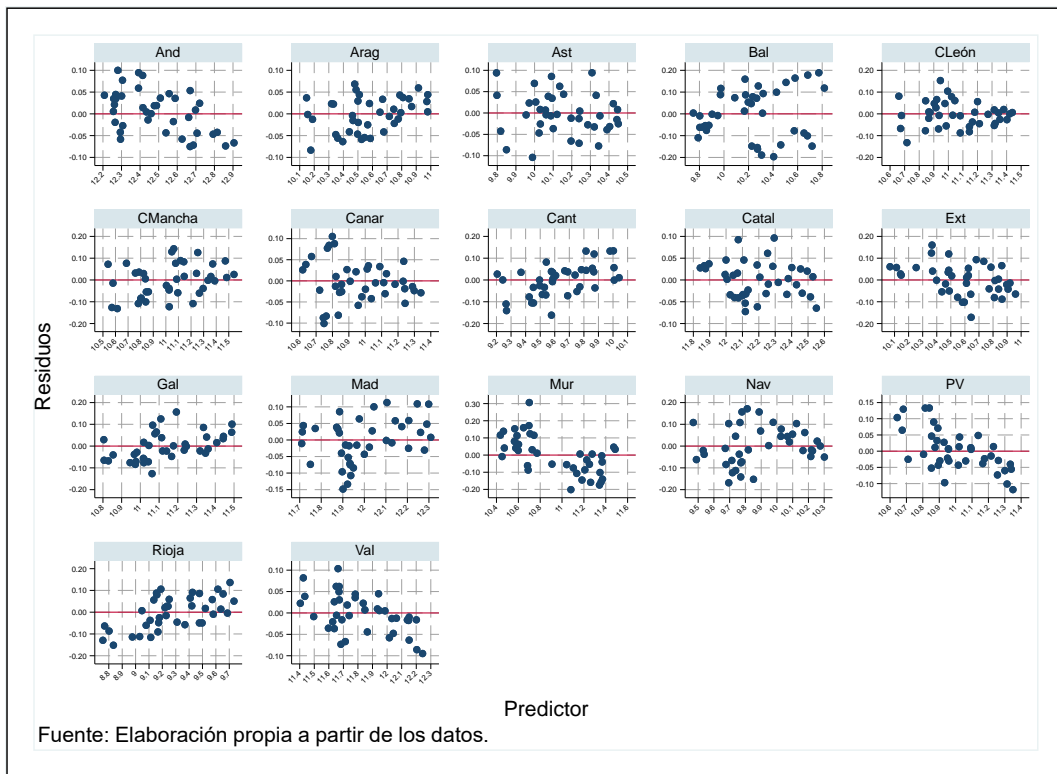


**Figura 5.2. Estimación de datos de panel aplicando el estimador de efectos fijos. Periodo 2009:T1-2018:T4.**

Como se puede observar, las series estimadas de contratos iniciales mediante el modelo de efectos fijos mejoran considerablemente a las estimadas con el panel agrupado (figura 5.1), siguiendo las series estimadas una tendencia y estacionalidad similar o próxima a las series reales y, por tanto, disminuyendo el término de residuo del modelo. Aplicando el estimador de efectos fijos se produce un re-escalamiento de las constantes de cada región

originando un mejor ajuste en cada una de ellas, reduciendo la diferencia entre ambas series, la real y la estimada. Este efecto individual o idiosincrático sobre la constante de cada unidad muestral se aprecia claramente en el caso de Andalucía, donde la serie de colocaciones predicha ha incrementado su intercepto notablemente al estimar por efectos fijos, mejorando la predicción dentro del periodo muestral.

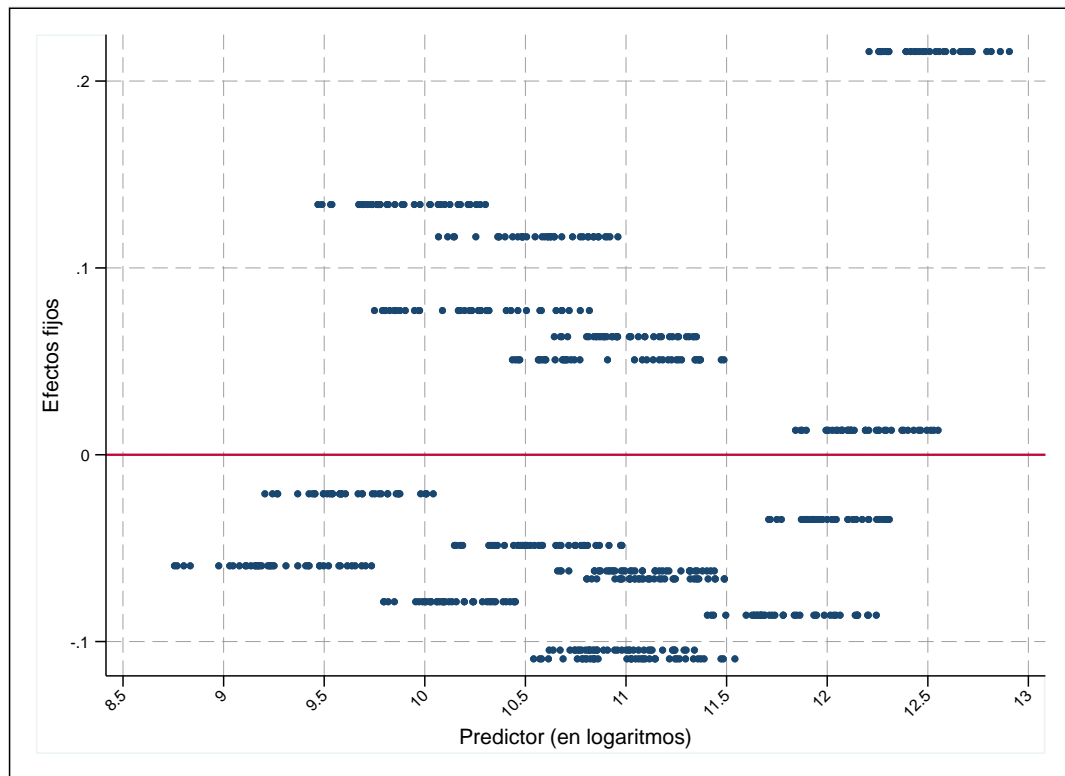
Nuestro análisis econométrico concluye evaluando la correcta especificación del modelo propuesto. En ocasiones, los modelos de regresión, como el panel de datos, pueden no estar especificados correctamente, bien sea por la omisión de variables relevantes o la saturación de un modelo al incluir variables innecesarias; ambas situaciones conducen al problema de error de especificación del modelo que acaba generando sesgos en los coeficientes estimados. Para determinar la fiabilidad de la estimación propuesta se suele recurrir, como primer paso, a un análisis de gráfico de diagnóstico. En concreto, es habitual enfrentar gráficamente a los residuos (diferencia entre la serie real y la estimada a partir del modelo) con el predictor –en este proyecto, el predictor consiste en la estimación de los flujos de colocaciones ( $M$ ) por CCAA obtenidos del modelo de la función de emparejamiento representado en la ecuación (5.1). La gráfica 5.3 representa dicha relación a nivel autonómico; si el modelo está bien estimado, los valores de los residuos se deben mover en torno a cero mostrando una varianza constante.



**Figura 5.3. Residuos de las series reales y las series estimadas frente al predictor. Periodo 2009:T1-2018:T4.**

Analizando la figura 5.3, se deduce que los residuos no siguen un patrón definido; mantienen un comportamiento aleatorio, independiente al nivel de predicción considerado. Aun así, existen excepciones como en la región de Murcia, que en niveles elevados de la predicción muestra residuos negativos, mientras que en niveles inferiores del predictor muestra residuos que toman, por lo general, valores mayores que cero. Por otro lado, ciertas comunidades se caracterizan por un patrón de residuos de pendiente positiva, obsérvese por ejemplo la Rioja. En cualquier caso, las singularidades del término de error descritas en determinadas comunidades requieren de una investigación más exhaustiva que escapa al alcance de este Trabajo.

Finalmente, el modelo de efectos fijos nos permite analizar gráficamente si existe algún indicio de correlación entre el predictor y los efectos individuales, tal y como se muestra en la figura 5.4. Recuérdese que si dichos elementos están correlacionados, el modelo de efectos aleatorios deja de ser consistente y, por tanto, lo consisten es estimar el modelo de efectos fijos.



**Figura 5.4. Efectos fijos frente al flujo de colocaciones o predictor en escala logarítmica (base 10). Periodo 2009:T1-2018:T4.**

La figura muestra que en niveles grandes de predicción los efectos fijos alcanzan valores pequeños, presentando una correlación negativa entre ambas variables. Esta estimación supone un indicio de correlación entre los efectos fijos y las variables explicativas ( $X$ ) y, por tanto, se debe estimar de forma consistente a través del modelo de efectos fijos, tal y como se ha deducido previamente, a partir del test de Hausman.

## 6. CONCLUSIONES.

El presente Trabajo de Fin de Máster se centra en explicar el comportamiento del emparejamiento entre trabajadores y vacantes en el mercado de trabajo español desde una perspectiva regional y sectorial. A nivel regional, desagregamos el análisis en las 17 Comunidades Autónomas que componen el territorio español; a nivel sectorial, empleamos información sobre la estructura sectorial de la generación de empleo de cada Comunidad. Para analizar el proceso de emparejamiento laboral a nivel regional desde un punto de vista econométrico (paramétrico), la metodología aplicada ha sido el Panel de Datos, que permite observar un número de individuos o unidades muestrales en diversos instantes de tiempo equidistantes entre sí, de modo que las unidades muestrales pueden ser seguidas en todos los periodos temporales –disponemos de un panel de datos balanceado.

Diversas resultan las teorías que intentan explicar la existencia de desequilibrio en el emparejamiento laboral y, por consiguiente, del desempleo continuo. La teoría no competitiva de búsqueda y emparejamiento desarrollada por Pissarides –véase Pissarides (2000)–, es el fundamento teórico del presente Trabajo de Fin de Máster, basada en una función de emparejamiento agregada. La función descrita, en su expresión más simple, muestra los emparejamientos ( $M$ ) formados en un determinado periodo de tiempo, a partir de los desempleados ( $U$ ) y los puestos de trabajo ( $V$ ) existentes durante dicho periodo.

Las series estadísticas registradas en los diversos organismos públicos, principalmente, la *Estadística de Empleo* del Servicio Público de Empleo Estatal (SEPE), han servido de base empírica del Trabajo, permitiendo generar y estimar el panel de datos a partir de observaciones por CCAA de las diferentes variables laborales empleadas durante el periodo de tiempo 2006-2018 –los datos mensuales del SEPE han sido trimestralizados para poder combinarlo con la estadística trimestral de vacantes que publica la *Encuesta de Coyuntura Laboral* (Ministerio de Trabajo, Migraciones y Seguridad Social) y la *Encuesta Trimestral del Coste Laboral* (INE). El modelo de panel de datos ha sido estimado mediante efectos fijos, el panel de datos tiene la ventaja, frente a los modelos solo de corte transversal o solo de corte temporal, de que permite controlar la heterogeneidad no solo observable sino también inobservable de cada unidad muestral o región.

Debido al ámbito temporal considerado, se hace diferenciación a lo largo del análisis entre los dos subperiodos bien diferenciados; la etapa de recesión experimentada durante los años 2009-2013 y la fase de expansión económica, la cual surge a comienzos de 2014 y dura hasta 2018 (el año 2019 da síntomas de cierto frenazo de la economía y no se considera en

este Trabajo). A nivel estadístico, las anomalías detectadas en las series de datos del SEPE durante los años siguientes al cambio que se produjo en el año 2005 en sus sistemas de información, aconsejan que nuestras estimaciones tomen a 2009 como primer año del periodo analizado.

Las principales conclusiones de nuestro análisis de panel de datos son tres. Por un lado, las variables explicativas del flujo de colocaciones no siempre muestran el signo esperado en la estimación del modelo. Por ejemplo, mientras que el signo del stock de desempleados es positivo y cercano a 0.2 (valor plausible dentro de la literatura en este campo), el flujo de nuevos desempleados (nuevos demandantes del SEPE) muestra un efecto negativo sobre el flujo de nuevos contratos que no se corresponde con lo que predice la teoría (el modelo *stock-flow* del emparejamiento laboral). En nuestra opinión, este signo negativo se puede deber a dos causas: (1) a que el carácter fuertemente anti-cíclico del nuevo desempleo está afectando al signo del coeficiente –hay que tener en cuenta que cuando la economía entra en recesión el flujo de nuevos parados aumenta rápidamente y de forma significativa–, (2) a que la variable de demandantes no resulta adecuada (representativa) para medir el volumen de buscadores de empleo, siendo más representativas series como la de los parados registrados o la de los DENOS publicadas por el SEPE, o como la serie de desempleados publicada en la EPA (INE).

Por otro lado, los coeficientes correspondientes a las series de “vacantes al comienzo del trimestre” y de “nuevas vacantes a lo largo del trimestre” tienen el signo positivo esperado (más vacantes permiten generar más contratos), pero muestran unos valores demasiado pequeños (no superan el valor 0,1 en ningún caso) si los comparamos con lo observado en otros trabajos sobre la economía española (como, por ejemplo, Álvarez de Toledo *et al.*, 2008). En efecto, los coeficientes de las variables de vacantes deberían tener valores más elevados, ya que en una economía como la española donde la demanda de trabajo por parte de las empresas es escasa en relación a la oferta de trabajo que aportan los trabajadores, el emparejamiento debería ser guiado fundamentalmente por el nivel de vacantes existente y, en especial, por las nuevas vacantes que fueran llegando al mercado cada trimestre. Desde nuestro punto de vista, el escaso valor de los coeficientes de las vacantes se puede deber a dos causas: (1) a que las series de vacantes de la economía española presentan problemas de representatividad, y (2) a que existe un fuerte desajuste o *mismatch* entre la oferta y la demanda de trabajo en la economía española; por ejemplo, a menudo se comenta entre las empresas españolas la dificultad que tienen para cubrir puestos *STEM* (*Science, Technology, Engineering, Math*).

Finalmente, nuestro Trabajo arroja luz sobre el efecto en la generación de empleo de ciertas condiciones del mercado de trabajo; esto es, condiciones que permiten generar un mayor volumen de colocaciones dado un nivel determinado de buscadores de empleo y de puestos vacantes. Así, aquellas regiones que muestran un porcentaje más elevado de contratación temporal generan un mayor volumen de colocaciones. Este dato no tiene por qué ser necesariamente bueno, ya que la mayor generación de empleo puede estar vinculada a la existencia de empleo de baja calidad. Por otro lado, las regiones con más peso de la agricultura y del sector servicios parecen haber soportado mejor el periodo de crisis (2009-2012) en términos de generación de empleo, mientras que en la época expansiva (2013-2018) tener un sector industrial importante contribuye a la mayor generación relativa de empleo.

Queda mucho por investigar sobre el proceso de emparejamiento laboral agregado en la economía española. Desde nuestro punto de vista, este Trabajo Fin de Máster invita a continuar la investigación en este campo abordando fundamentalmente dos temas, la mejora de las estadísticas sobre el emparejamiento laboral (sobre todo en lo que se refiere a la información sobre el total de vacantes que hay en el territorio español en cada periodo), y la propuesta de nuevos modelos teóricos y econométricos que permitan capturar y controlar estadísticamente la heterogeneidad existente entre los trabajadores, por un lado, y entre las vacantes, por otro.



## 7. REFERENCIA BIBLIOGRÁFICA.

Álvarez de Toledo, P., Núñez, F., Usabiaga, C. (2008). *La Función de Emparejamiento en el Mercado de Trabajo Español*. Revista de Economía Aplicada, 16 (48), 5-35.

Antolín, P. (1994). *Unemployment flows and vacancies in Spain*. IVIE, Documento de Trabajo (serie EC) n.º 9405.

Baum, C.F. (2006). *An Introduction to Modern Econometrics Using Stata*. Stata Press: United States of America.

Blanchard, O.J. y P. Diamond (1994). *Ranking, Unemployment Duration and Wages*. The Review of Economics Studies, 61(3), 417-434.

Burda, M.C. y C.Wyplosz (1994). *Gross Worker and Job Flows in Europe*. European economic review, 38 (6), 1287-1315.

Burdett, K., M.G. Coles y J.C. Van Ours (1994). *Temporal Aggregation Bias in Stock-Flow Models*. CEPR Discussion Paper, 967.

Coles, M.G. (1994). *Understanding the Matching Function: The Role of Newspapers and Job Agencies*. CEPR Discussion Paper, 939.

Coles, M.G. y E. Smith (1996). *Cross-Section Estimation of the Matching Function: Evidence from England and Wales*. Economica, 63 (252), 589-598.

Coles, M.G. y E. Smith (1998). *Marketplaces and Matching*. International Economic Review, 39 (1), 239-254.

Gujarati, D.N. y D.C. Porter (2010). *Econometría*. México: McGraw-Hill.

Instituto Nacional de Estadística. *Encuesta Trimestral de Coste Laboral*. [https://www.ine.es/prensa/etcl\\_prensa.htm](https://www.ine.es/prensa/etcl_prensa.htm)

Kennedy, P.E. (2008). *A Guide to Econometrics*. Estados Unidos: The MIT Press.

Layard, R., S.J. Nickell y R.A. Jackman (1991). *Unemployment, Macro-economic Performance and the Labour Market*. Oxford: Oxford University Press.

Ministerio de Trabajo, Migraciones y Seguridad Social. *Encuesta de Coyuntura Laboral*. <http://www.mitramiss.gob.es/estadisticas/ECL/welcome.htm>

Ministerio de Trabajo, Migraciones y Seguridad Social. *Servicio Público de Empleo Estatal*. <https://www.google.com/search?q=sepe&oq=sepe+&aqs=chrome..69i57j0l2j69i60l3.2046j0j4&sourceid=chrome&ie=UTF-8>

Münich, D., J.Svejnar y K.Terrel (1999). *Worker-Firm Matching and Unemployment in Transition to a Market: (Why) Were the Czechs More Successful than Other*. Working paper, 107, Davidson Institute, University of Michigan Business School.

Nuñez, F. y Usabiaga, C. (2007). *La curva de Beveridge y la Función de Emparejamiento: Revisión de sus Fundamentos Teóricos y de la Literatura Empírica, con Especial Énfasis en el Caso Español*. España: Centro de Estudios Andaluces.

Petrongolo, B. y C.A. Pissarides (2001). *Looking into the Black Box: A survey of the Matching Function*. *Journal of Economic Literature*, 39 (2), 390-431.

Petrongolo, B. y Coles, Melvyn (2003). *A test between unemployment theories using matching data*. CEPR Discussion Paper, 273.

Pissarides, C.A. (1986). Unemployment and Vacancies in Britain. *Economic policy*, 1 (3), 676-690.

Pissarides, C.A. (2000). *Equilibrium Unemployment Theory*. Inglaterra: The MIT Press.

Shimer, R. y Smith, L. (2001). *Matching, Search, and Heterogeneity*. *Advances in Macroeconomics*, 1 (1), 1010-1029.