

¿Quién se empareja con quién en el mercado laboral español? Un análisis *cluster* basado en la *Muestra Continua de Vidas Laborales*

Pablo Álvarez de Toledo Saavedra^a ▪ Fernando Núñez Hernández^a
▪ Carlos Usabiaga Ibáñez^{b,c}

Resumen

Nuestro trabajo plantea herramientas empíricas que sirven para captar el papel de las heterogeneidades en el proceso de emparejamiento laboral. Desarrollamos una aplicación de nuestra metodología al mercado de trabajo español basada en los datos de colocaciones contenidos en la *Muestra Continua de Vidas Laborales* (MCVL). Estos datos contienen información sobre las características del trabajador y del puesto vacante que forman cada colocación, lo cual nos permite realizar una agrupación inicial de unos y otros en segmentos de trabajador y de puesto en función de dichas características. Estos segmentos pueden, a su vez, ser agrupados respectivamente en *clusters* que, finalmente, pueden ser combinados entre sí según las colocaciones que se hayan producido entre ellos. Aquellos *clusters* conjuntos (o *biclusters*) con una mayor propensión al emparejamiento pueden ser considerados mercados generadores de empleo, en cuya estructura

Manuscrito recibido el 3 de marzo de 2016; aceptado el 19 de septiembre de 2016.

- a Departamento de Organización Industrial y Gestión de Empresas I de la Escuela Superior de Ingenieros de la Universidad de Sevilla (España).
- b Departamento de Economía, Métodos Cuantitativos e Historia Económica de la Universidad Pablo de Olavide (España).
- c Autor para correspondencia: cusaiba@upo.es.

Agradecemos las sugerencias recibidas de José María Arranz y de los evaluadores y editores de la revista. Este trabajo ha sido presentado en las *IX Jornadas de Economía Laboral* (Barcelona, 2015); en el *INFER Workshop on Applied Macroeconomics and Labour Economics* (Sevilla, 2015); en el *I Workshop de la Asociación Andaluza de Ciencia Regional* (Sevilla, 2015); en el *XIX Encuentro de Economía Aplicada* (Sevilla, 2016), y en seminarios en la Universidad Loyola-Andalucía, la Universidad Pablo de Olavide y la Universidad de Sevilla. En todos esos foros hemos recibido valiosas sugerencias. También agradecemos la financiación recibida de los siguientes Proyectos de Investigación: ECO2012-35430 (Ministerio de Economía y Competitividad español) y PRY 055/12 y SEJ-513 (Junta de Andalucía).

© 2017 Universidad Nacional Autónoma de México, Facultad de Economía. Este es un artículo *Open Access* bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

podemos profundizar —por ejemplo, analizando la duración del empleo—. Nuestra metodología empírica, que puede ser muy versátil en su aplicación (de lo que ofrecemos varios ejemplos), sirve para visualizar la importante segmentación del mercado de trabajo y puede ser útil para el diseño eficiente de políticas activas de empleo, como las de movilidad laboral.

Palabras clave: emparejamiento laboral, segmentación, análisis *cluster*, políticas activas del mercado de trabajo, información sobre el empleo, *Muestra Continua de Vidas Laborales*.

Clasificación JEL: J61, J62, J63, J64.

Abstract

Our study addresses empirical tools that are useful to capture the role of heterogeneities in the labor matching process. We develop an application of our methodology to the Spanish labor market based on the data from the *Muestra Continua de Vidas Laborales* (*Continuous Sample of Working Lives*). These data contain information about the worker and the job that form each job placement, which allows us to group in worker and job segments on the basis of their characteristics. These segments can be also grouped respectively in worker and job clusters, which finally can be combined among them attending to their matching. Those joint clusters (or biclusters) with the highest propensity to match can be considered specific “job creation” markets, whose structure can be studied in detail—for example, analyzing the employment duration—. Our empirical methodology, which can be very versatile in its application (of which we provide several examples), allows showing the important labor market segmentation, and it can be useful for the efficient design of active employment policies, for instance of labor mobility.

Keywords: Labor matching; segmentation; cluster analysis; active labor market policies; information on employment; *Continuous Sample of Working Lives*.

JEL Classification: J61, J62, J63, J64.

INTRODUCCIÓN

Una posible estrategia para analizar bases de datos que contienen una gran cantidad de observaciones o elementos consiste en clasificarlos en un número determinado de *clusters* o grupos, de manera tal que cada agrupación presente, en un escenario ideal, una alta homogeneidad interior y una alta heterogeneidad respecto a otras agrupaciones. Así, más que analizar las características de cada elemento de la muestra por separado, lo que hacemos es analizar las características de los grupos de esos elementos que se han formado.¹

1 Kaufman y Rousseeuw (2005) comienzan su conocido texto indicando: “Cluster analysis is the art of finding groups in data”. Otros trabajos sobre el análisis *cluster* que pueden consultarse son Cotterman y Peracchi (1992), que proponen una clasificación industrial; el *survey* de Jain, Murty y Flynn (1999), y el

Existen multitud de técnicas para agrupar o clasificar datos en *clusters*; todas ellas han de hacer frente al menos a tres cuestiones metodológicas: generar una medida de proximidad o similitud entre los elementos que se van a agrupar, definir un procedimiento jerarquizado o no de formación de los grupos y proponer una regla de parada o de determinación de grupos en el proceso de *clustering*.

Álvarez de Toledo, Núñez y Usabiaga (2014) proponen una medida de similitud que tiene como principal característica su aplicabilidad a datos de emparejamiento oferta-demanda; por ejemplo, dos demandantes (o dos grupos reducidos de ellos) —empresas que demandan trabajadores, familias que buscan vivienda, ahorradores que demandan activos financieros, etc.— serán tanto más parecidos cuanto más similar sea la forma en que se emparejan con el otro lado del mercado. Este análisis se puede aplicar a los dos lados del mercado, agrupando en *clusters* a oferentes por un lado y a demandantes por otro, siendo posible, además, obtener un *bicluster* del mercado mediante la combinación de los dos *clusters* anteriores con base en los intercambios o emparejamientos producidos entre ellos. El presente trabajo aplica esta metodología de *bicluster* al mercado de trabajo español,² y más concretamente a la información sobre altas laborales (colocaciones) disponible en la *Muestra Continua de Vidas Laborales* (MCVL) durante el periodo 2011-2013.³ Dicha metodología permite, a partir de la información disponible para cada colocación, dividir el mercado de trabajo en *clusters* (o mercados de trabajo específicos) a través del agrupamiento, no de trabajadores o puestos individuales, sino de una serie de pequeñas agrupaciones de ellos defi-

texto introductorio de Everitt *et al.* (2011). El análisis *cluster* es también una herramienta muy utilizada dentro de las técnicas de minería de datos y de *big data*.

- 2 Es bien conocido que el mercado de trabajo español es problemático, especialmente en ciertas regiones, como Andalucía (García-Cintado, Romero-Ávila y Usabiaga, 2014). Por ejemplo, podemos recordar la elevada y persistente tasa de desempleo, la débil trayectoria de la productividad y el excesivo peso de la contratación temporal. Entre las explicaciones subyacentes a esos problemas podemos citar, entre otros muchos factores barajados, la problemática *insider-outsider*, los factores vinculados al desempleo de larga duración, las rigideces salariales, el modelo de negociación colectiva —que presenta un grado de centralización intermedio—, las modalidades de contratación (debate sobre el “contrato único”) y la reducida movilidad geográfica.
- 3 Se trata de una encuesta española muy importante. Por ello, hemos encontrado numerosos trabajos que tratan de explotar la información de la MCVL, por ejemplo desde la perspectiva del flujo de entrada al empleo y desde la perspectiva del cobro de prestaciones —véanse como muestra los trabajos de Cebrián y Toharia (2008), Lapuerta (2010), Cebrián, Moreno y Toharia (2011) y Arranz y García-Serrano (2011, 2014)—, pero ninguno de los trabajos consultados aplica un análisis *cluster* o similar a dicha base de datos. En este sentido, pensamos que la combinación de esta técnica con esta encuesta puede ser muy relevante.

nidas según las características del trabajador que se emplea en el momento de su búsqueda (características como el sexo, la edad, el municipio de residencia, el grupo de ocupación y el sector de actividad) y según las características del puesto vacante que se ocupa (municipio del centro de trabajo, grupo de ocupación y sector de actividad de la vacante). Dichas agrupaciones de partida suponen la división del mercado de trabajo en segmentos laborales de trabajador y de puesto; nuestro análisis se centra en dichos segmentos laborales más que en los trabajadores y en los puestos considerados individualmente.

Esta segmentación de partida del mercado de trabajo se ajusta bien a la forma en que describen el intercambio laboral los modelos de emparejamiento con fricciones y búsqueda por los dos lados del mercado —en esta línea, véanse, por ejemplo, los trabajos de Pissarides (2000, 2011), Petrongolo y Pissarides (2001) y Shimer (2007)—. En estos modelos, la búsqueda de empleo que subyace al desempleo no consiste únicamente en encontrar un puesto con buen salario, sino en hallar un buen emparejamiento en su conjunto. Por otro lado, no sólo es el trabajador el que se preocupa por encontrar un buen puesto, con la empresa simplemente dispuesta a contratar a aquél que acepte su oferta salarial, sino que la empresa también está preocupada por la localización de un buen candidato. La base del modelo de emparejamiento es que cada trabajador tiene características distintas, que le convierte en adecuado o no según el tipo de trabajo. Asimismo, los requisitos de cada puesto de trabajo difieren entre las empresas, y los empleadores no son indiferentes sobre el tipo de trabajador que contratan, dado un nivel salarial. Este enfoque microeconómico de los dos lados del emparejamiento laboral permite incorporar características del mundo real, como las diferencias que existen entre los trabajadores y entre los puestos de trabajo, o las diferencias existentes en la estructura institucional de los mercados de trabajo, que pueden ayudar a la interpretación de los resultados observados; por ejemplo, respecto al comportamiento del desempleo.

Dentro de los modelos microeconómicos que se están desarrollando en este campo, nuestro trabajo puede enlazar al menos con dos tipos de modelos: los que tratan de microfundamentar la función de emparejamiento y los de asignación bilateral (*two-sided matching games*). Los primeros pretenden hacer explícitas las fricciones y heterogeneidades que subyacen en dicha función de emparejamiento agregada.⁴ Así, modelos como el de islas, bolas en urnas, taxis,

4 Esta función, que relaciona básicamente las colocaciones con los puestos vacantes y los desempleados, asume de forma implícita la existencia de heterogeneidades, fricciones e imperfecciones de la

colas, *stock-flow* o el de desajuste o *mismatch*, han explorado diferentes tipos de fricciones que permiten microfundamentar a la función de emparejamiento, dividiendo de una forma u otra el mercado de trabajo en diferentes partes o segmentos.⁵ En cuanto a los segundos (los modelos de *two-sided matching games*), pueden también constituir un marco teórico adecuado para nuestro trabajo, ya que analizamos datos individuales de colocaciones detectando “quién se empareja con quién” (*who matches with whom*); colocaciones que se producen, según este enfoque, a partir de una determinada función o tecnología de asignación bilateral basada en las preferencias de los individuos.⁶

Nuestro trabajo no tiene la intención de ampliar o evaluar los modelos teóricos comentados, más bien intenta proponer unas variables empíricas que pueden tener implicaciones importantes en el marco de esos modelos. Por ejemplo, hay que tener en cuenta que el “mapa” que generamos de propensiones al emparejamiento (entre segmentos de trabajadores y segmentos de puestos), y la posibilidad de identificar segmentos (de trabajadores o de puestos) con una elevada similitud en cuanto a la forma en que participan en el emparejamiento laboral, pueden ser interpretados como procesos que mejoran la información sobre el empleo; mejoras que pueden producir una reducción del *mismatch* existente en la función de emparejamiento o en la tecnología subyacente a los modelos de asignación bilateral —pudiendo incluso provocar una revisión de las preferencias de los individuos en dichos “juegos” de asignación—.

Este trabajo, al analizar *bichusters* o mercados de trabajo específicos reales de la economía española, profundiza y amplía el análisis propuesto por Álvarez de Toledo, Núñez y Usabiaga (2014).⁷ Este último trabajo plantea la posibi-

información en el emparejamiento laboral, así como la idea de que dicho proceso se efectúa con costos de búsqueda. La función agregada de emparejamiento no asume que los trabajadores y los puestos de trabajo son homogéneos, simplemente omite hacer explícitas las heterogeneidades existentes.

- 5 Sobre la función de emparejamiento y su fundamento, pueden consultarse los trabajos de Petrongolo y Pissarides (2001) y Núñez y Usabiaga (2007) —este último trabajo centrado en aplicaciones para el caso español—. Por ejemplo, sobre evidencia del tipo *stock-flow* para la economía española, véanse Álvarez de Toledo, Núñez y Usabiaga (2008, 2011).
- 6 Los modelos de asignación bilateral del emparejamiento se aplicaron inicialmente al “mercado” matrimonial —véanse, por ejemplo, los trabajos de Gale y Shapley (1962), Becker (1973) y Fox (2010)—. Algunas aplicaciones al mercado de trabajo de esta metodología se pueden encontrar en Boyd *et al.* (2013) y Agarwal (2015).
- 7 Véase también en esta línea el trabajo previo, más descriptivo, de Álvarez de Toledo, Núñez y Usabiaga (2013).

lidad de agrupar el mercado laboral en mercados locales o “islas” laborales, pero lo hace usando una medida de similitud común a ambos lados del mercado —mientras que en este trabajo analizamos las similitudes por separado— y, además, no llega a implementar empíricamente dicha clasificación.⁸ En suma, este estudio supone un importante paso adelante desde varias perspectivas (metodología, tratamiento de los datos, etc.) y podría ser útil como herramienta para mejorar el emparejamiento y la política laboral, como se ejemplifica para la economía española.

El resto del trabajo se estructura de la siguiente forma. En la sección segunda explicamos la metodología seguida para segmentar una base de datos sobre emparejamientos —dando lugar a una matriz de emparejamientos entre segmentos de trabajador y segmentos de puesto— y generar las variables empíricas que sinteticen su distribución; variables como la similitud o proximidad entre dos segmentos laborales. En la sección tercera procesamos y agrupamos los datos de colocaciones contenidos en la MCVL con el objetivo de generar la matriz de emparejamientos entre segmentos de trabajador y segmentos de puesto; matriz cuya dimensión y estructura deben ser operativas computacionalmente para poder obtener las variables descritas en la sección anterior y para poder generar, a partir de dichas variables, un “mapa” *bicluster*. Por su parte, en la sección cuarta se muestra cómo la información sobre el flujo de entrada al empleo se puede sintetizar dando lugar a la generación de una serie de mercados laborales idiosincrásicos (*biclusters*), cuya estructura puede ser analizada también desde el punto de vista de la duración del empleo (sección quinta). Finalmente, la sección sexta recoge las principales conclusiones de nuestro trabajo.

8 También existen diferencias entre nuestro trabajo y el de Álvarez de Toledo, Núñez y Usabiaga (2014) en lo que se refiere a los datos empleados. En primer lugar, hay que tener en cuenta que Álvarez de Toledo, Núñez y Usabiaga (2014) emplean datos de la economía andaluza procedentes de las oficinas públicas de empleo durante el periodo 2007-2010, mientras que el presente trabajo utiliza los datos españoles procedentes de la MCVL durante el periodo 2011-2013. En segundo lugar, nuestro trabajo trata de obtener, mediante el agrupamiento de segmentos laborales, un mapa de la segmentación del mercado de trabajo español en su totalidad; por su parte, Álvarez de Toledo, Núñez y Usabiaga (2014), al considerar únicamente segmentos laborales caracterizados por generar un volumen elevado de colocaciones, analizan sólo una parte del mercado de trabajo.

METODOLOGÍA

Este trabajo toma como base metodológica el esquema empírico de segmentación propuesto por Álvarez de Toledo, Núñez y Usabiaga (2014). Dicho documento define, a partir de una base de datos de colocaciones, una serie de segmentos laborales según las características del trabajador que se emplea (características correspondientes al periodo de búsqueda del individuo, como el sexo, la edad, el municipio de residencia, la ocupación y el sector de actividad) y según las características del puesto vacante que se ocupa (municipio del centro de trabajo, ocupación y sector de actividad de la vacante). En cada periodo t cada trabajador que busca un empleo es asignado a uno de los n *segmentos de trabajador* existentes, W_{it} ($i = 1, 2, \dots, n$); cada vacante es asignada a uno de los m *segmentos de puesto*, J_{jt} ($j = 1, 2, \dots, m$), y cada uno de los emparejamientos trabajador-puesto da lugar a un *segmento conjunto*, S_{ijt} , formado por la unión del segmento del trabajador y del puesto. Cada tipo de segmento viene definido por el correspondiente conjunto de características $\bar{W}_i, \bar{J}_j, \bar{S}_{ij}$, donde \bar{S}_{ij} incluye las características del segmento de trabajador \bar{W}_i y del segmento de puesto \bar{J}_j que se emparejan.⁹

En cada periodo t dispondremos de una matriz $n \times m$ de distribución de frecuencias o colocaciones, donde las filas representan a los segmentos de trabajador y las columnas a los segmentos de puesto. Cada elemento de la matriz M_{ijt} representa el número de emparejamientos que se ha producido en el segmento conjunto correspondiente; la matriz completa muestra, por tanto, quién se empareja con quién.¹⁰ El número total de colocaciones, $M_t = \sum_{i=1}^n \sum_{j=1}^m M_{ijt}$, es la suma de todos los elementos de la matriz, es decir, la suma de las frecuencias o colocaciones observadas en todos los segmentos conjuntos, y la frecuencia o cantidad de emparejamientos en cada segmento de trabajador y en cada segmento de puesto viene dada, respectivamente, por la suma en la matriz de la fila y de la columna correspondiente: $M_{it} = \sum_{j=1}^m M_{ijt}$, $M_{jt} = \sum_{i=1}^n M_{ijt}$.

⁹ Barnichon y Figura (2015) usan formalmente un tipo similar de segmentación laboral.

¹⁰ En los juegos de asignación bilateral del emparejamiento, el “quién se empareja con quién” viene determinado por una función tecnológica de emparejamiento basada en preferencias —véase, por ejemplo, Fox (2008)—.

Interpretaremos que el mercado de trabajo está segmentado cuando los trabajadores de un segmento de trabajador específico tienden a emparejarse con las vacantes de un determinado segmento de puesto en mayor o menor medida a como lo harían si la asignación del empleo en el mercado fuera puramente aleatoria. En relación con esta idea de segmentación, proponemos dos medidas empíricas obtenidas a partir de nuestros datos de emparejamiento: la “propensión al emparejamiento” entre un segmento de trabajador y un segmento de puesto y el “grado de similitud” entre dos segmentos de trabajador (o dos segmentos de puesto).

La propensión al emparejamiento entre un segmento de trabajador i y un segmento de puesto j determinados en cada periodo t , pm_{ijt} , se obtiene como el cociente entre la probabilidad, de acuerdo con nuestros datos, de observar una colocación en el segmento conjunto que forman y la probabilidad de dicho emparejamiento en el caso de que el proceso de emparejamiento sea puramente aleatorio (colocaciones asignadas por “sorteo”).

$$pm_{ijt} = \frac{p_{ijt}}{\widehat{p_{ijt}}} = \frac{p_{ijt}}{p_{it}p_{jt}} = \frac{\frac{M_{ijt}}{M_t}}{\frac{M_{it}}{M_t} \frac{M_{jt}}{M_t}} = M_t \frac{M_{ijt}}{M_{it}M_{jt}} \quad [1]$$

Debido a que M_{ijt} es menor o igual que M_{it} y que M_{jt} , el valor de pm_{ijt} oscilará entre cero y M_t —valdrá cero cuando $M_{ijt} = 0$ y tomará un valor máximo de M_t en el caso particular de que $M_{ijt} = M_{it} = M_{jt} = 1$ —. Por otro lado, una propensión mayor a la unidad indicaría que la “conexión” entre el segmento de trabajador y el segmento de puesto es más fuerte que la que se obtendría en un escenario de asignación aleatoria de las colocaciones, y a la inversa. Según el enfoque *two-sided*, una propensión alta entre trabajadores de W_i y puestos de J_j se deberá a la combinación de tres factores: preferencia de los trabajadores tipo i por los puestos tipo j ; preferencia de los puestos tipo j (nos referimos a las personas que hacen la selección para esos puestos) por los trabajadores tipo i , y empleo de canales de búsqueda comunes.

En cuanto a la medida de similitud, consideramos a dos segmentos de trabajador, W_{i1} y W_{i2} , “similares” cuando sus emparejamientos se distribuyen porcentualmente de forma parecida entre los distintos segmentos de puesto —esta medida de similitud puede ser expresada en función de la variable pm_{ij} —:

$$sw_{i1-i2t} = \sum_{j=1}^m \min\left(\frac{M_{i1jt}}{M_{i1t}}, \frac{M_{i2jt}}{M_{i2t}}\right) = \sum_{j=1}^m \min(pm_{i1jt}, pm_{i2jt})p_{jt} \quad [2]$$

El valor de la similitud, sw_{i1-i2t} , oscila entre cero y uno. Cuando es cero, W_{i1} no dirige ninguna colocación a los segmentos de puesto a los que dirige sus colocaciones W_{i2} . Cuando es uno, las distribuciones de las colocaciones de ambos segmentos de trabajador son idénticas. Análogamente, consideramos a dos segmentos de puesto “similares” cuando sus emparejamientos se distribuyen de forma parecida entre los distintos segmentos de trabajador.

GENERACIÓN DE DATOS LABORALES SEGMENTADOS

A PARTIR DE LA MCVL

Nuestro trabajo aborda la aplicación de una metodología de *clusters* a la información disponible en la MCVL de 2013 sobre las altas laborales producidas durante el periodo 2011-2013. Por lo tanto, el objetivo es analizar al colectivo de personas que entra a formar parte de la muestra porque ha transitado al empleo en algún momento dentro del periodo, dando lugar al alta laboral correspondiente en la Seguridad Social; y esto con independencia de la situación administrativa de perceptor o no de prestación o subsidio que tenga el trabajador empleado —además, hay que tener en cuenta que un individuo puede ser al mismo tiempo afiliado en alta y pensionista, entrando así en la población de referencia de la muestra por ambas vías—.

A partir de estos datos de emparejamientos, podemos dividir el mercado de trabajo en segmentos de trabajador y en segmentos de puesto atendiendo a las características o atributos de los trabajadores y de los puestos. En concreto, nos centraremos en cinco características del trabajador (en el momento de su búsqueda) que pueden influir en su emparejamiento: municipio, grupo de ocupación, tramo de edad, sexo y sector de actividad; y en tres características del puesto cubierto: municipio, grupo de ocupación y sector de actividad. Así, cada colocación dará lugar a un segmento conjunto formado por ocho características, las cinco del segmento de trabajador y las tres del segmento del puesto. El siguiente cuadro resume las variables de la MCVL que hemos empleado para obtener cada característica.

Cuadro 1
Variables empleadas procedentes de la MCVL

		Características	Variable MCVL	Fichero MCVL
Segmento conjunto	Segmento de trabajador	Sexo	Sexo	Personas
		Edad por tramos	Edad = Fecha real del alta en afiliación (1) – Fecha de nacimiento (2)	(1) Afiliación y (2) Personas
		Municipio	Domicilio de residencia habitual	Personas
		Grupo de ocupación	Grupo de cotización (del empleo anterior)	Afiliación
		Sector de actividad	Actividad económica de la cuenta de cotización (CNAE 2009) (del empleo anterior)	Afiliación
	Segmento de puesto	Municipio	Domicilio de actividad de la cuenta de cotización	Afiliación
		Grupo de ocupación	Grupo de cotización (del empleo actual)	Afiliación
		Sector de actividad	Actividad económica de la cuenta de cotización (CNAE 2009) (del empleo actual)	Afiliación

La variable edad, que se corresponde con la edad (en años) que tiene el trabajador en el momento en que consigue una colocación, se ha expresado en intervalos para evitar una excesiva segmentación de nuestros datos. De esta forma, los intervalos que hemos considerado han sido: 16 a 29 años; 30 a 44 años; 45 a 54 años, y más de 54 años. En cuanto a los municipios (del trabajador y del puesto), hay que indicar que, lamentablemente, no es posible identificar en la muestra a los municipios con población igual o inferior a 40 000 habitantes, los cuales son agrupados por provincia. Cabe señalar también que nos hemos aproximado al atributo *grupo de ocupación* a través de la variable *grupo de cotización*. Finalmente, nuestro análisis supone que el grupo de ocupación del trabajador y su sector de actividad (durante la búsqueda de empleo) son el grupo de ocupación y el sector de actividad del último empleo que ha tenido —evidentemente, si se trata del primer empleo, estas variables han quedado sin informar—. Pensamos que todos estos supuestos o criterios son plausibles y no distorsionan apenas los resultados obtenidos.

Los datos utilizados representan algo más de dos millones de colocaciones o altas laborales. De todas estas colocaciones, hemos descartado aquellas para las que no se conoce el municipio o la ocupación del trabajador o del puesto,

características esenciales a la hora de introducir segmentación en el mercado de trabajo.¹¹ Tras aplicar dicha restricción, quedan algo menos de dos millones de colocaciones (1 967 441),¹² las cuales dan lugar a 585 420 segmentos conjuntos no nulos (segmentos donde se ha observado al menos una colocación), 139 504 segmentos de trabajador y 39 981 segmentos de puesto. Por consiguiente, según la estructura de nuestros datos, el mercado laboral español está claramente segmentado, es decir, existe un número reducido de segmentos conjuntos no nulos, que muestran por lo general propensiones al emparejamiento elevadas, y una gran cantidad de segmentos conjuntos que muestran frecuencias conjuntas nulas y, por ende, propensiones al emparejamiento nulas.¹³ Sin embargo, hay que tener en cuenta que nuestros datos (a pesar de su importante volumen) podrían estar ofreciendo una visión del mercado de trabajo demasiado segmentada. Pensemos que si dispusiéramos de una muestra de colocaciones significativamente más grande que la empleada (MCVL), se podrían observar colocaciones en un mayor número de segmentos conjuntos, existiendo más propensiones al emparejamiento mayores que cero. Este problema de carácter muestral se conoce con el nombre de “problema de las frecuencias nulas”.¹⁴ Además, no sólo el problema de las frecuencias nulas puede condicionar nuestros resultados, sino también la existencia de frecuencias demasiado bajas.¹⁵ Para corregir estos problemas hemos adoptado dos soluciones que pretenden conseguir, de forma acumulativa, una matriz de emparejamientos con mayores

11 En nuestra base de datos, las variables edad y sexo del trabajador están siempre informadas, pero no sucede lo mismo para las variables municipio, grupo de ocupación y sector de actividad.

12 El 59.1% de estas colocaciones tiene una duración menor a un mes y el 74.5% una duración inferior a tres meses. Por lo tanto, la mayor parte del empleo generado tiene carácter temporal. En este trabajo optamos por incluir el empleo de carácter temporal en todo el proceso de segmentación y de *clustering* propuesto, de manera que no se desaproveche ninguna información sobre el empleo. A este respecto, nuestro estudio también aporta un análisis de las duraciones del empleo en los diferentes *biclusters* obtenidos.

13 Álvarez de Toledo, Núñez y Usabiaga (2014) proponen una medida del grado de segmentación o aislamiento de un segmento laboral determinado. Estos autores encuentran un elevado grado de segmentación laboral en su análisis del mercado de trabajo andaluz.

14 El problema de las frecuencias nulas puede ser tratado con técnicas de alisamiento (Laplace, *add-one smoothing*, etc.). Véase, a este respecto, el trabajo de Liu (2011).

15 Segmentos conjuntos con frecuencias relativamente bajas pueden mostrar valores de pm_{ij} demasiado elevados. Por ejemplo, en el caso de que un segmento de trabajador y un segmento de puesto formen una única colocación entre ellos y no formen colocaciones con ningún otro segmento ($M_{ij} = 1$, $M_i = 1$ y $M_j = 1$), ello da lugar a una pm_{ij} máxima que puede ser discutible.

frecuencias. Por un lado, hemos considerado los tres años analizados (2011 a 2013)¹⁶ como un periodo único; ello implica que renunciamos a adoptar un enfoque dinámico, que considere cada año por separado, a cambio de poder acumular un mayor número de colocaciones observadas en cada segmento. Por otro lado, hemos optado por unir entre sí a segmentos de trabajador con frecuencias reducidas (inferiores a diez colocaciones) y, a su vez, a éstos con segmentos de mayor frecuencia. Dicho proceso de unión ha tenido un elevado carácter endógeno, ya que se ha basado fundamentalmente en la similitud entre los municipios de cada par de segmentos —cuanto más se parecen en la forma en que se emparejan con los diferentes segmentos de puesto, más similares son dos municipios de trabajador—, sujeto a la coincidencia del resto de sus características. Todo este proceso de agregación se ha llevado también a cabo con los segmentos de puesto. Tras estos agrupamientos, el número de segmentos de trabajador ha quedado reducido a 9 170 (segmentos que engloban a los 139 504 de partida), y el número de segmentos de puesto ha quedado reducido a 9 208 (segmentos que engloban a los 39 981 de partida).¹⁷ Hay que tener en cuenta que las sucesivas agrupaciones de segmentos que llevamos a cabo en este trabajo no suponen pérdida alguna de información, esto es, en cada momento conocemos los segmentos iniciales u originales que hay dentro de cada agrupación y las frecuencias o colocaciones de éstos; de hecho, el análisis descriptivo de cada *bicluster* laboral se basará en dichos segmentos de partida.

ANÁLISIS BICLUSTER DEL MERCADO DE TRABAJO ESPAÑOL

En esta sección, los algo más de 9 000 segmentos de trabajador y segmentos de puesto obtenidos en la sección anterior serán agrupados en *clusters* jerárquicos, *clusters* que serán posteriormente combinados en función de los emparejamientos

16 En los años analizados (2011 a 2013), la crisis económica afecta muy gravemente al mercado laboral español. Así, según los datos de la *Encuesta de Población Activa* del Instituto Nacional de Estadística, el número total de ocupados pasó de aproximadamente 18.4 a 17.1 millones. Asimismo, en ese periodo bajó la tasa de actividad y subieron la tasa de desempleo, el peso del desempleo de larga duración y el peso del empleo a tiempo parcial. Incluso en este difícil o crítico contexto laboral, donde se llegaron a superar los seis millones de desempleados, nuestra metodología muestra que puede ser una herramienta útil para identificar focos de generación de empleo.

17 Los algo más de 9 000 segmentos (de trabajador o de puesto) son el resultado de la unión de dos grupos de segmentos: los de mayor frecuencia, que no han entrado en la agrupación, y los restantes, que han sido objeto de agrupamientos.

para generar *biclusters*¹⁸ que permitan obtener una visión panorámica de la segmentación existente en el mercado de trabajo español y realizar otros análisis.

El proceso comienza con la obtención de un *cluster* jerárquico de segmentos de trabajador. Es decir, vamos a ir agrupando los 9 170 segmentos de trabajador de la etapa anterior en grupos o *clusters* de segmentos, basándonos en la medida de similitud entre ellos ($sw_{i_1-i_2}$) recogida en la ecuación [2]. En esencia, consideramos a dos segmentos de trabajador “similares” cuando sus emparejamientos se distribuyen de forma muy parecida entre los distintos segmentos de puesto. A partir de aquí, seguimos un método jerárquico, en el que se van agrupando progresivamente los segmentos menos distantes (más próximos) en *clusters*, y los *clusters* entre sí para formar *clusters* mayores (que contienen cada vez un mayor número de segmentos). Esto es, se comienza considerando cada segmento por separado, después se agrupan los dos segmentos entre los que existe mayor proximidad (menor distancia), y así sucesivamente, pudiendo continuarse el proceso hasta la fusión del total en un único segmento de trabajador.

Hay que tener en cuenta que cada ronda de agrupación requeriría actualizar la matriz de emparejamientos —al tener que acumularse las frecuencias de los elementos (segmentos o *clusters*) que se han unido— y volver a calcular las similitudes de todos los elementos existentes de cara a una siguiente ronda de agrupamiento. Sin embargo, existen buenas aproximaciones que pueden evitar este recálculo continuo, como el procedimiento de “enlace promedio” que finalmente hemos empleado, en el que en cada ronda de agrupamiento la distancia entre dos *clusters* se calcula como la media de todas las distancias de los elementos de uno y otro *cluster*.

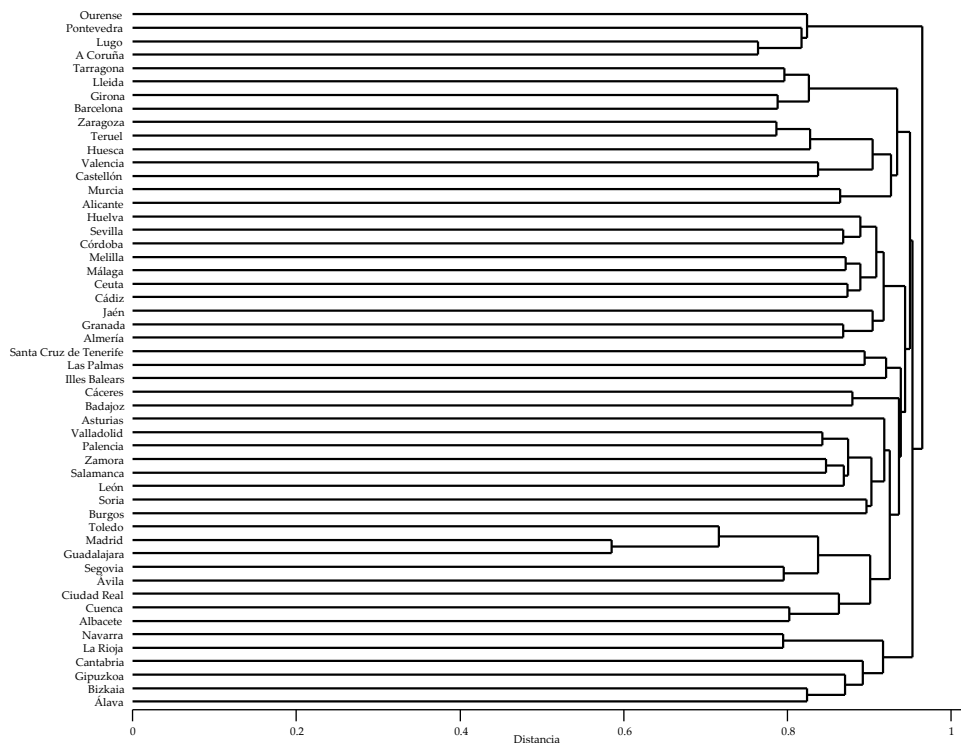
El proceso de agrupamiento se puede detener en un número de *clusters* determinado previamente o, por ejemplo, tratando de garantizar la existencia de *clusters* cuyos elementos no superen una determinada distancia. Además, los resultados del proceso pueden mostrarse gráficamente mediante un *dendrograma*, figura en forma de árbol en la que se observa cómo se van formando *clusters* cada vez mayores y más distantes entre sí.

A modo de ejemplo, las siguientes figuras muestran los respectivos dendrogramas aplicados a las provincias, a los grupos de ocupación y a los sectores de actividad de los trabajadores, considerando cada variable por separado. Las

18 El término *bicluster* procede del campo de la investigación celular —véase, por ejemplo, Panteli, Boutsinas y Giannikos (2014)—.

provincias de trabajador se van agrupando entre sí por lo parecidas que son en la forma en que se emparejan con las diferentes provincias de puesto, y análogamente se realiza para los grupos de ocupación y sectores de actividad. Estas figuras ofrecen una información preliminar de interés.

Figura 1
Dendrograma de las provincias de trabajador

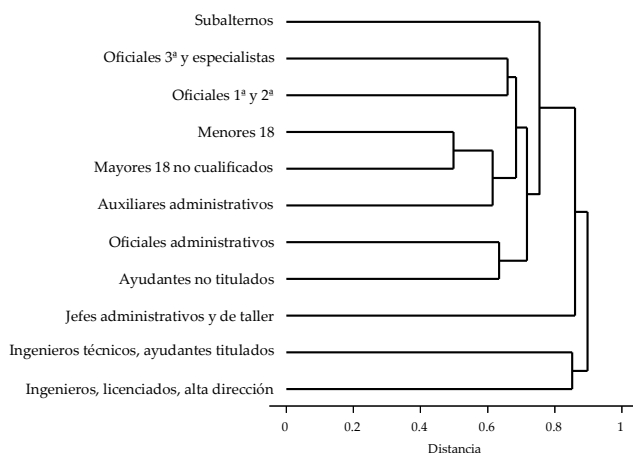


El dendrograma de las provincias de trabajador (figura 1) muestra que las dos primeras provincias en formar un *cluster* son Madrid y Guadalajara. A continuación, dicho *cluster* se une con Toledo para formar un nuevo *cluster*, de tres provincias. La siguiente unión de provincias es la que se produce entre Lugo y A Coruña. Se puede apreciar que estas primeras provincias apuntadas son cercanas geográficamente. El proceso de fusiones continuaría hasta que todas las provincias quedasen agrupadas en un único *cluster*.

Por su parte, el dendrograma por grupos de ocupación (figura 2) muestra un hecho interesante: los trabajadores cualificados pertenecientes al grupo de

“ingenieros, licenciados y alta dirección” no se parecen demasiado en el emparejamiento a los trabajadores, también cualificados, del grupo “ingenieros técnicos y ayudantes titulados”. Se puede observar en la figura que los grupos que se unen primero (más próximos desde la perspectiva del emparejamiento) corresponden a trabajadores de baja cualificación.

Figura 2
Dendrograma de las ocupaciones de trabajador

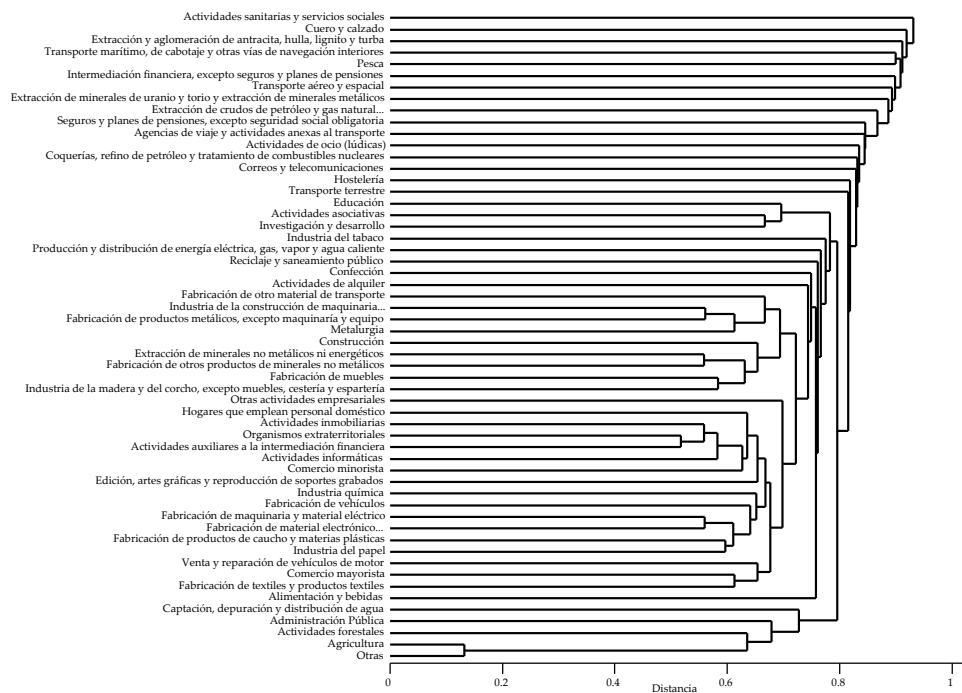


Finalmente, el dendrograma por sector de actividad (figura 3) muestra cómo se van agrupando, de forma progresiva, las actividades relacionadas entre sí. Ciertas actividades, como las sanitarias o las relacionadas con el cuero, la minería o el transporte marítimo, aparecen bastante diferenciadas de las demás, es decir, sus trabajadores se emparejan de una forma específica, que tiene poco que ver con la forma en la que se emparejan los trabajadores de otros sectores de actividad.

Volviendo al conjunto de los 9 170 segmentos de trabajador, lo que proponemos es analizar el dendrograma de *clusters* de segmentos de trabajador cuando restan 256 *clusters*, dado que hemos comprobado que para dicho nivel de agrupamiento empiezan a detectarse suficientes *clusters* con un número aceptable de elementos o segmentos de trabajador.¹⁹ El proceso de agrupamiento hasta

¹⁹ A medida que un *cluster* va creciendo, con los sucesivos agrupamientos, pierde homogeneidad entre sus elementos. Cuando tenemos 256 *clusters* de segmentos de trabajador, la disimilitud existente entre los diferentes segmentos de cada *cluster* ronda en promedio el 75% —dato que ha sido calculado

Figura 3
Dendrograma de las actividades de trabajador



256 *clusters* se ha aplicado también a los 9 208 segmentos de puesto. En este caso, la medida de similitud entre segmentos de puesto se basa en la idea de que dos segmentos de puesto serán tanto más parecidos cuando más se parezca la forma en que sus emparejamientos se distribuyen entre los distintos segmentos de trabajador.

Una vez que tenemos los 256 *clusters* de segmentos de trabajador y los 256 de segmentos de puesto, podemos proceder a combinarlos entre sí, dando lugar a un *bicluster* del mercado de trabajo que se puede representar fundamentalmente a través de dos matrices de dimensiones 256×256: la de frecuencias o colocaciones, y la de propensiones al emparejamiento entre *clusters*. Ambas matrices están formadas por 256 filas, que representan a los *clusters* de segmentos de trabajador, y 256 columnas, que representan a los *clusters* de segmentos de puesto. Cada uno

con aquellos *clusters* que tienen al menos dos segmentos—; a pesar de esta cifra relativamente alta, hemos de tener en cuenta que dentro de cada *cluster* puede haber subgrupos de segmentos con una disimilitud (dentro del subgrupo) bastante más reducida.

de los 65 536 elementos (o *biclusters*) de la matriz de frecuencias representa la acumulación de las colocaciones de los segmentos conjuntos contenidos en el cruce o *bicluster* correspondiente, esto es, la acumulación de las colocaciones que se han formado entre los segmentos del correspondiente *cluster* de trabajador y los segmentos del correspondiente *cluster* de puesto—. Por su parte, la matriz de propensiones representa la propensión al emparejamiento entre cada *cluster* de trabajador y cada *cluster* de puesto correspondiente; calculada dicha propensión a partir de las colocaciones acumuladas en la matriz de frecuencias anteriormente descrita y de acuerdo con la ecuación [1].

Un *bicluster* con una propensión al emparejamiento nula es un *bicluster* donde los segmentos del *cluster* de trabajador no se relacionan con los segmentos del *cluster* de puesto. Por el contrario, un *bicluster* que muestre una propensión al emparejamiento relativamente elevada señala que los trabajadores de los segmentos de trabajador del *bicluster* muestran una elevada propensión a emparejarse con los puestos vacantes de los segmentos de puesto del *bicluster*, y a la inversa; este segundo escenario puede ser considerado como un yacimiento de empleo o mercado generador de empleo “específico o idiosincrásico”.²⁰

Hay que tener en cuenta que, de los 65 536 elementos de la matriz de propensiones, se han observado valores positivos en 5 072 casos, es decir, hemos identificado 5 072 *biclusters* o mercados laborales específicos, cuyos principales parámetros se recogen en el siguiente cuadro.

Cuadro 2
Características del *bicluster*

	<i>Propensión al emparejamiento</i>	<i>Segmentos conjuntos dentro del bicluster</i>	<i>Colocaciones del bicluster</i>	<i>Colocaciones del cluster de trabajador</i>	<i>Colocaciones del cluster de puesto</i>
Media	65	115	388	41 351	24 082
Percentil 50	0.2	4	7	17 906	9 502
Mínimo	0.001	1	1	10	10
Máximo	19 640	46 666	136 028	238 553	151 258

²⁰ Conviene puntualizar que nuestro análisis, al centrarse en las colocaciones que se producen en el mercado de trabajo, sólo permite observar a trabajadores que han accedido a un empleo y a puestos que han sido cubiertos, de manera que, siendo precisos, más que tratar de identificar mercados de trabajo locales en un sentido amplio (considerando toda la oferta y la demanda de trabajo), se trata de identificar focos (o *biclusters*) de generación de empleo.

Llegados a este punto, puede resultar ilustrativo realizar un análisis descriptivo de aquellos *biclusters* que puedan suscitar interés por cumplir ciertas propiedades. En este sentido, el cuadro 3 muestra la estructura de cinco mercados laborales (*biclusters*) que se caracterizan por cumplir, cada uno de ellos, un triple criterio: el *cluster* de trabajador presenta una frecuencia (o cantidad de colocaciones) relativamente elevada; el *cluster* de puesto muestra una frecuencia relativamente

Cuadro 3
Bicluster del mercado laboral español. Elevadas frecuencias de trabajador y de puesto y alta propensión al emparejamiento

Cluster de segmentos de trabajador	Cluster de segmentos de puesto	Propensión emparejamiento	Segmentos conjuntos en el bicluster	Colocaciones cluster segmentos trabajador	Colocaciones cluster segmentos puesto	Colocaciones bicluster	Sexo	Sexo (%)	Edad	Edad (%)
75	164	7.4	46 666	238 553	151 258	136 028	Hombres	55.4	30-44	44.9
							Mujeres	44.6	16-29	36.3
									45-54	14.8
									55-	3.9
41	73	7.6	37 606	220 995	138 945	118 349	Hombres	58.7	30-44	43.6
							Mujeres	41.3	16-29	31.0
									45-54	18.3
									55-	7.1
21	206	9.7	16 617	182 520	94 863	85 484	Hombres	56.5	30-44	45.5
							Mujeres	43.5	16-29	27.5
									45-54	20.4
									55-	6.6
36	117	15.0	10 983	120 493	58 785	54 040	Hombres	62.3	30-44	49.7
							Mujeres	37.7	16-29	29.8
									45-54	15.7
									55-	4.8
30	101	14.7	24 444	112 233	108 831	91 397	Hombres	58.3	30-44	47.1
							Mujeres	41.7	16-29	31.2
									45-54	16.4
									55-	5.3

elevada al compararlo con otros *clusters* de puesto que forman un *bicluster* con el *cluster* de trabajador seleccionado; y la propensión al emparejamiento entre ambos *clusters* es elevada. Tanto por el lado de los trabajadores como por el de los puestos, a cada *cluster* se le asigna un número identificativo (recogido en las dos primeras columnas del cuadro) que es relevante dentro del proceso de agrupamiento jerárquico, ya que se refiere al orden de los elementos de partida del *cluster* dentro del dendrograma.

<i>Provincia trabajador</i>	<i>Provincia trabajador (%)</i>	<i>Provincia puesto</i>	<i>Provincia puesto (%)</i>	<i>Municipio trabajador</i>	<i>Municipio trabajador (%)</i>	<i>Municipio puesto</i>	<i>Municipio puesto (%)</i>
Barcelona	92.2	Barcelona	95.6	Municipio ≤ 40 000 Barcelona	28.6	Barcelona	54.0
Girona	4.0	Girona	2.7	Barcelona	26.7	Municipio ≤ 40 000 Barcelona	18.9
Tarragona	2.7	Tarragona	1.1	L'Hospitalet de Llobregat	5.2	Sant Cugat del Vallès	3.6
Lleida	0.3	Lleida	0.3	Badalona	3.9	L'Hospitalet de Llobregat	3.1
Huesca	0.2	Huesca	0.2	Terrassa	3.3	Municipio ≤ 40 000 Girona	2.0
Otros	0.5	Otros	0.2	Otros	32.3	Otros	18.5
Madrid	78.3	Madrid	79.8	Madrid	34.6	Madrid	51.1
Toledo	14.2	Toledo	13.2	Municipio ≤ 40 000 Toledo	11.6	Municipio ≤ 40 000 Toledo	8.7
Guadalajara	5.5	Guadalajara	5.3	Municipio ≤ 40 000 Madrid	8.7	Municipio ≤ 40 000 Madrid	6.0
Ávila	0.8	Ávila	0.8	Alcalá de Henares	4.5	Coslada	5.1
Segovia	0.6	Segovia	0.7	Fuenlabrada	3.8	Toledo	3.7
Otros	0.7	Otros	0.3	Otros	36.9	Otros	25.3
Sevilla	93.2	Sevilla	97.2	Municipio ≤ 40 000 Sevilla	62.4	Municipio ≤ 40 000 Sevilla	53.8
Córdoba	3.6	Córdoba	1.3	Sevilla	18.8	Sevilla	31.0
Huelva	2.1	Cádiz	0.9	Dos Hermanas	4.6	Dos Hermanas	3.6
Cádiz	1.0	Huelva	0.4	Utrera	2.7	Écija	2.8
Illes Balears	0.1	Badajoz	0.1	Municipio ≤ 40 000 Córdoba	2.5	Utrera	2.6
Otros	0.0	Otros	0.1	Otros	9.0	Otros	6.2
Murcia	89.7	Murcia	91.7	Municipio ≤ 40 000 Murcia	46.0	Municipio ≤ 40 000 Murcia	42.6
Alicante	9.6	Alicante	7.7	Murcia	19.3	Murcia	26.1
Madrid	0.4	Almería	0.6	Cartagena	12.1	Cartagena	13.8
Albacete	0.2			Lorca	7.5	Municipio ≤ 40 000 Alicante	5.8
Almería	0.1			Municipio ≤ 40 000 Alicante	5.8	Lorca	5.8
Otros	0.0			Otros	9.4	Otros	6.0
Valencia	82.3	Valencia	81.6	Municipio ≤ 40 000 Valencia	47.0	Municipio ≤ 40 000 Valencia	36.9
Castellón	17.3	Castellón	17.8	Valencia	23.3	Valencia	34.9
Granada	0.2	Lleida	0.2	Municipio ≤ 40 000 Castellón	11.9	Municipio	11.3
Alicante	0.1	Albacete	0.2	Castellón de la Plana	4.0	Castellón de la Plana	5.1
Madrid	0.0	Huesca	0.1	Gandía	3.7	Paterna	3.5
Otros	0.0	Otros	0.1	Otros	10.2	Otros	8.1

Cuadro 3, continuación...

Cluster de segmentos de trabajador	Cluster de segmentos de puesto	Ocupación trabajador	Ocupación trabajador (%)	Ocupación puesto	Ocupación puesto (%)	Actividad trabajador	Actividad trabajador (%)	Actividad puesto	Actividad puesto (%)
75	164	Peones y asimilados	32.2	Peones y asimilados	32.9	Otras actividades empresariales	42.0	Otras actividades empresariales	43.4
		Oficiales 1ª y 2ª	20.8	Oficiales 1ª y 2ª	21.1	Hostelería	16.1	Hostelería	16.6
		Oficiales 3ª y especialistas	18.4	Oficiales 3ª y especialistas	18.6	Actividades de ocio	7.6	Actividades de ocio	7.7
		Auxiliares administrativos	15.7	Auxiliares administrativos	15.6	Construcción	7.4	Comercio minorista	7.5
		Oficiales administrativos	7.2	Oficiales administrativos	7.7	Comercio minorista	7.0	Construcción	6.9
		Otros	5.7	Otros	4.1	Otros	19.9	Otros	17.9
41	73	Peones y asimilados	38.0	Peones y asimilados	41.0	Otras actividades empresariales	40.6	Otras actividades empresariales	42.0
		Oficiales 1ª y 2ª	28.5	Oficiales 1ª y 2ª	29.7	Hostelería	21.5	Hostelería	23.3
		Oficiales 3ª y especialistas	17.3	Oficiales 3ª y especialistas	18.1	Construcción	11.2	Construcción	10.9
		Auxiliares administrativos	6.2	Subalternos	6.8	Comercio minorista	3.3	Transporte terrestre	2.9
		Subalternos	5.9	Auxiliares administrativos	2.3	Transporte terrestre	2.7	Agricultura	2.2
		Otros	4.1	Otros	2.1	Otros	20.8	Otros	18.7
21	206	Peones y asimilados	51.7	Peones y asimilados	56.6	Agricultura	29.8	Agricultura	33.8
		Oficiales 1ª y 2ª	19.0	Oficiales 1ª y 2ª	19.1	Otras actividades empresariales	15.6	Otras actividades empresariales	15.9
		Auxiliares administrativos	13.5	Oficiales 3ª y especialistas	9.4	Construcción	12.1	Construcción	11.7
		Oficiales 3ª y especialistas	9.5	Auxiliares administrativos	9.3	Hostelería	10.1	Hostelería	10.5
		Oficiales administrativos	3.7	Oficiales administrativos	3.6	Administración Pública	4.5	Administración Pública	5.1
		Otros	2.6	Otros	2.0	Otros	27.8	Otros	23.0
36	117	Peones y asimilados	56.5	Peones y asimilados	64.1	Otras actividades empresariales	28.2	Otras actividades empresariales	30.7
		Oficiales 3ª y especialistas	15.0	Oficiales 3ª y especialistas	15.1	Agricultura	19.0	Agricultura	25.1
		Oficiales 1ª y 2ª	14.5	Oficiales 1ª y 2ª	14.1	Hostelería	17.7	Hostelería	18.4
		Auxiliares administrativos	10.3	Auxiliares administrativos	3.7	Construcción	6.6	Construcción	5.7
		Oficiales administrativos	1.4	Subalternos	1.0	Otras actividades	6.4	Alimentación y bebidas	4.9
		Otros	2.4	Otros	1.9	Otros	22.1	Otros	15.3
30	101	Peones y asimilados	37.4	Peones y asimilados	40.6	Otras actividades empresariales	32.3	Otras actividades empresariales	33.3
		Oficiales 1ª y 2ª	25.3	Oficiales 1ª y 2ª	24.3	Hostelería	19.4	Hostelería	20.0
		Oficiales 3ª y especialistas	18.1	Oficiales 3ª y especialistas	18.4	Construcción	6.5	Agricultura	8.0
		Auxiliares administrativos	11.3	Auxiliares administrativos	8.5	Comercio mayorista	6.4	Comercio mayorista	6.6
		Oficiales administrativos	4.1	Oficiales administrativos	3.8	Agricultura	6.1	Comercio minorista	5.6
				Otros	29.3	Otros	26.6		

Lo primero que podemos destacar del cuadro 3 es que todos los *clusters* de trabajador superan las 100 000 colocaciones en el periodo objeto de estudio, mientras que los *biclusters* y los *clusters* de puesto superan, todos, las 50 000 colocaciones. Se trata, por tanto, de yacimientos de empleo importantes. Además, la propensión a emparejarse supera el valor 7 en todos los casos —se trata de propensiones situadas en el cuartil más alto de la distribución de propensiones—.

En cuanto a las características de los trabajadores, observamos mayores porcentajes de hombres que de mujeres, especialmente en el *bicluster* 36-117, situado fundamentalmente en la provincia de Murcia. El tramo de edad mayoritario es el comprendido entre los 30 y 44 años, seguido de los trabajadores más jóvenes (16 a 29 años).

A nivel espacial, se observa que los trabajadores y los puestos de cada *bicluster* tienden a concentrarse en áreas geográficas bastante parecidas, destacando también el hecho de que los municipios con una población menor o igual a 40 000 habitantes de una provincia determinada, considerados conjuntamente, pueden llegar a aportar más trabajadores y/o más puestos que las propias capitales de provincia; esto sucede de forma más significativa en el *bicluster* 21-206, situado fundamentalmente en la provincia de Sevilla. Finalmente, se observa que los diferentes mercados identificados están dominados por trabajadores y puestos de baja o media cualificación, siendo los sectores de actividad predominantes los de otras actividades empresariales:²¹ hostelería, agricultura y construcción.

Dado que los mercados laborales comentados se construyen básicamente sobre trabajadores de cualificación baja o media, vamos a incluir en nuestro estudio el análisis de una serie de *biclusters* que se caracterizan por la presencia mayoritaria de trabajadores con estudios superiores —grupos de cotización 1 y 2— (véase el cuadro 4), siguiendo un criterio de selección de *biclusters* similar al utilizado en el caso anterior.

21 Este sector incluye actividades jurídicas y de contabilidad; actividades de consultoría de gestión empresarial; servicios técnicos de arquitectura e ingeniería; ensayos y análisis técnicos; publicidad y estudios de mercado; otras actividades profesionales, científicas y técnicas; actividades relacionadas con el empleo; actividades de seguridad e investigación; servicios a edificios y actividades de jardinería; actividades administrativas de oficina, y otras actividades auxiliares a las empresas.

Cuadro 4
Bicluster del mercado laboral español.
Trabajadores cualificados

<i>Cluster de segmentos de trabajador</i>	<i>Cluster de segmentos de puesto</i>	<i>Propensión emparejamiento</i>	<i>Segmentos conjuntos en el bicluster</i>	<i>Colocaciones cluster segmentos trabajador</i>	<i>Colocaciones cluster segmentos puesto</i>	<i>Colocaciones bicluster</i>	<i>Sexo</i>	<i>Sexo (%)</i>	<i>Edad</i>	<i>Edad (%)</i>
88	158	34.7	1 737	34 854	28 346	17 428	Mujeres	80.3	16-29	52.7
							Hombres	19.7	30-44	37.0
									45-54	6.9
									55-	3.4
54	90	60.1	600	25 552	12 626	9 857	Mujeres	84.1	16-29	52.0
							Hombres	15.9	30-44	42.7
									45-54	4.5
									55-	0.8
90	141	96.0	280	15 465	5 169	3 899	Mujeres	55.2	30-44	61.6
							Hombres	44.8	45-54	18.2
									16-29	11.5
									55-	8.7
9	18	109.3	317	13 906	4 831	3 732	Mujeres	79.8	30-44	49.2
							Hombres	20.2	16-29	34.7
									45-54	10.9
									55-	5.2
121	203	47.0	1 076	12 844	10 045	3 079	Mujeres	65.0	30-44	59.1
							Hombres	35.0	16-29	26.2
									45-54	12.2
									55-	2.5
112	30	169.6	142	11 309	6 015	5 865	Mujeres	91.5	16-29	49.4
							Hombres	8.5	30-44	45.8
									45-54	3.7
									55-	1.1

<i>Provincia trabajador</i>	<i>Provincia trabajador (%)</i>	<i>Provincia puesto</i>	<i>Provincia puesto (%)</i>	<i>Municipio trabajador</i>	<i>Municipio trabajador (%)</i>	<i>Municipio puesto</i>	<i>Municipio puesto (%)</i>
Barcelona	88.0	Barcelona	92.6	Barcelona	32.6	Barcelona	67.0
Girona-	5.1	Girona	4.0	Municipio ≤ 40 000 Barcelona	26.6	L'Hospitalet de Llobregat	7.6
Tarragona	4.0	Tarragona	3.3	L'Hospitalet de Llobregat	4.7	Municipio ≤ 40 000 Barcelona	5.0
Sevilla	0.9	Lleida	0.0	Municipio ≤ 40 000 Girona	3.6	Girona	2.5
Valencia	0.8	Teruel	0.0	Terrassa	3.1	Mataró	2.1
Otros	1.2			Otros	29.4	Otros	15.8
Madrid	59.6	Madrid	60.7	Madrid	22.1	Madrid	46.8
Toledo	11.6	Toledo	10.8	Municipio ≤ 40 000 Toledo	7.4	Segovia	8.3
Albacete	6.3	Segovia	8.3	Municipio ≤ 40 000 Madrid	7.2	Toledo	7.7
Segovia	6.1	Ciudad Real	6.7	Municipio ≤ 40 000 Segovia	5.4	Albacete	5.1
Cuenca	5.9	Albacete	6.4	Móstoles	5.0	Cuenca	4.6
Otros	10.4	Otros	7.2	Otros	52.8	Otros	27.4
Valencia	53.8	Valencia	54.8	Valencia	23.6	Valencia	52.5
Alicante	28.4	Alicante	27.3	Municipio ≤ 40 000 Valencia	19.5	Alicante	24.8
Castellón	12.8	Castellón	15.7	Alicante	13.1	Castellón de la Plana	15.7
Madrid	1.7	Murcia	2.2	Municipio ≤ 40 000 Alicante	10.6	Murcia	1.3
Teruel	1.1			Castellón de la Plana	6.2	Municipio ≤ 40 000 Valencia	1.3
Otros	2.0			Otros	27.0	Otros	4.4
Bizkaia	86.4	Bizkaia	89.1	Municipio ≤ 40 000 Bizkaia	38.1	Bilbao	47.2
Gipuzkoa	6.6	Gipuzkoa	6.8	Bilbao	29.3	Municipio ≤ 40 000 Bizkaia	23.1
Álava	3.4	Álava	3.4	Barakaldo	8.6	Barakaldo	16.1
Cantabria	2.2	Navarra	0.5	Getxo	4.5	San Sebastián	4.4
Málaga	0.8	La Rioja	0.1	San Sebastián	3.8	Vitoria	3.3
Otros	0.7			Otros	15.7	Otros	5.9
Sevilla	60.9	Sevilla	72.3	Municipio ≤ 40 000 Sevilla	26.7	Sevilla	54.1
Cádiz	20.7	Cádiz	19.5	Sevilla	25.8	Municipio ≤ 40 000 Sevilla	12.4
Córdoba	9.9	Córdoba	7.9	Municipio ≤ 40 000 Córdoba	5.7	Cádiz	9.4
Huelva	3.3	Huelva	0.4	Municipio ≤ 40 000 Cádiz	5.5	Municipio ≤ 40 000 Córdoba	3.9
Málaga	2.5			Córdoba	4.0	Córdoba	3.7
Otros	2.8			Otros	32.3	Otros	16.5
A Coruña	53.4	A Coruña	58.0	Municipio ≤ 40 000 A Coruña	28.2	Vigo	31.6
Pontevedra	45.2	Pontevedra	42.0	Municipio ≤ 40 000 Pontevedra	24.7	Santiago de Compostela	27.1
Ourense	1.0			Vigo	17.5	A Coruña	18.1
Lugo	0.4			A Coruña	13.1	Ferrol	10.2
				Santiago de Compostela	8.8	Pontevedra	8.8
				Otros	7.6	Otros	4.2

Cuadro 4, continuación...

Cluster de segmentos de trabajador	Cluster de segmentos de puesto	Ocupación trabajador	Ocupación trabajador (%)	Ocupación puesto	Ocupación puesto (%)
88	158	Ingenieros técnicos, ayudantes titulados	93.5	Ingenieros técnicos, ayudantes titulados	92.5
		Ingenieros, licenciados, alta dirección	6.3	Ingenieros, licenciados, alta dirección	5.3
		Oficiales administrativos	0.1	Ayudantes no titulados	0.8
		Auxiliares administrativos	0.1	Oficiales administrativos	0.5
		Jefes administrativos y de taller	0.0	Jefes administrativos y de taller	0.3
		Otros	0.0	Otros	0.6
54	90	Ingenieros técnicos, ayudantes titulados	99.7	Ingenieros técnicos, ayudantes titulados	100.0
		Ingenieros, licenciados, alta dirección	0.3		
		Ayudantes no titulados	0.0		
		Jefes administrativos y de taller	0.0		
90	141	Ingenieros, licenciados, alta dirección	99.6	Ingenieros, licenciados, alta dirección	100.0
		Ingenieros técnicos, ayudantes titulados	0.4		
9	18	Ingenieros técnicos, ayudantes titulados	71.7	Ingenieros técnicos, ayudantes titulados	71.4
		Ingenieros, licenciados, alta dirección	28.3	Ingenieros, licenciados, alta dirección	28.6
		Subalternos	0.0	Jefes administrativos y de taller	0.1
121	203	Ingenieros técnicos, ayudantes titulados	100.0	Ingenieros técnicos, ayudantes titulados	91.3
				Ingenieros, licenciados, alta dirección	3.1
				Jefes administrativos y de taller	2.3
				Auxiliares administrativos	1.0
				Oficiales administrativos	0.6
		Otros	1.7		
112	30	Ingenieros técnicos, ayudantes titulados	100.0	Ingenieros técnicos, ayudantes titulados	100.0

En conjunto, y a diferencia de los mercados anteriormente comentados, el cuadro muestra que las colocaciones observadas en los diferentes *biclusters* corresponden principalmente a mujeres. En cuanto a la edad del trabajador en el momento de la colocación, predominan los tramos de 30 a 44 años y de 16 a 29 años; en los *biclusters* situados fundamentalmente en las provincias de Madrid y Barcelona

<i>Actividad trabajador</i>	<i>Actividad trabajador (%)</i>	<i>Actividad puesto</i>	<i>Actividad puesto (%)</i>
Actividades sanitarias y servicios sociales	70.6	Actividades sanitarias y servicios sociales	70.4
Administración Pública	18.4	Administración Pública	17.5
Educación	7.7	Educación	9.3
Investigación y desarrollo	1.1	Investigación y desarrollo	1.4
Actividades asociativas	1.0	Actividades asociativas	1.2
Otros	1.2	Otros	0.2
Actividades sanitarias y servicios sociales	99.0	Actividades sanitarias y servicios sociales	100.0
Administración Pública	0.5		
Educación	0.5		
Otras actividades empresariales	0.0		
Actividades de ocio	0.0		
Otros	0.0		
Actividades sanitarias y servicios sociales	96.5	Actividades sanitarias y servicios sociales	98.3
Comercio minorista	1.7	Comercio minorista	1.7
Educación	0.6		
Investigación y desarrollo	0.4		
Administración Pública	0.3		
Otros	0.5		
Actividades sanitarias y servicios sociales	97.2	Actividades sanitarias y servicios sociales	99.0
Actividades asociativas	1.2	Actividades asociativas	1.0
Educación	0.5		
Administración Pública	0.3		
Otras actividades empresariales	0.3		
Otros	0.5		
Actividades sanitarias y servicios sociales	33.2	Actividades sanitarias y servicios sociales	31.5
Educación	29.4	Educación	30.4
Administración Pública	10.7	Administración Pública	11.7
Otras actividades empresariales	6.5	Otras actividades empresariales	5.6
Actividades asociativas	5.3	Actividades asociativas	5.3
Otros	14.9	Otros	15.5
Actividades sanitarias y servicios sociales	97.2	Actividades sanitarias y servicios sociales	98.0
Actividades asociativas	2.1	Actividades asociativas	2.0
Administración Pública	0.4		
Educación	0.1		
Otras actividades empresariales	0.1		
Otros	0.1		

tienen un mayor peso los trabajadores más jóvenes, sobre todo en el caso de Barcelona.

A nivel espacial, la muestra de *biclusters* de trabajadores cualificados se sitúa en torno a las provincias de Barcelona, Madrid, Valencia, Bizkaia, Sevilla y Pontevedra, lo cual sucede tanto por el lado de los trabajadores como por el de los

puestos, es decir, el *cluster* de trabajador y el de puesto se sitúan en zonas geográficas parecidas. Destaca, además, el hecho, a nivel del análisis de los municipios, de que en los *clusters* de puesto los municipios que son capital de provincia tienen mayor peso que en los *clusters* de trabajador, esto es, las capitales ofrecen un volumen relativamente importante de puestos, los cuales son ocupados por trabajadores que proceden de la propia capital o de municipios cercanos.

En cuanto a los grupos de ocupación, dado nuestro criterio adicional de selección, el cuadro 4 ofrece *biclusters* donde los trabajadores pertenecen mayoritaria-

Cuadro 5
Bicluster del mercado laboral español.
Desajuste o mismatch en el emparejamiento

Cluster de segmentos de trabajador	Cluster de segmentos de puesto	Propensión emparejamiento	Segmentos conjuntos en el bicluster	Colocaciones cluster segmentos trabajador	Colocaciones cluster segmentos puesto	Colocaciones bicluster	Sexo	Sexo (%)	Eddad	Eddad (%)	Provincia trabajador	Provincia trabajador (%)
99	35	16.4	180	72 224	743	447	Hombres	82.1	45-54	42.1	A Coruña	63.1
							Mujeres	17.9	30-44	37.1	Pontevedra	33.3
									55-16-29	13.4	Lugo	3.4
									7.4	Ourense	0.2	
96	140	2.4	23	3 346	15 410	62	Mujeres	62.9	30-44	66.1	Las Palmas	67.7
							Hombres	37.1	55-	19.4	Santa Cruz de Tenerife	32.3
									45-54	12.9		
									1.6			
112	33	1.0	302	11 309	60 750	353	Mujeres	64.3	30-44	55.2	Pontevedra	39.1
							Hombres	35.7	16-29	32.3	A Coruña	36.3
									45-54	9.6	Ourense	14.4
									2.8	Lugo	10.2	
9	4	1.2	304	13 906	41 814	340	Mujeres	56.2	30-44	49.4	Bizkaia	52.4
							Hombres	43.8	16-29	40.3	Álava	22.4
									45-54	8.8	Gipuzkoa	16.2
									1.5	La Rioja	4.7	
										Navarra	4.1	
										Otros	0.3	

mente a los grupos de cotización 1 (“ingenieros, licenciados y alta dirección”) y 2 (“ingenieros técnicos, peritos y ayudantes”). En general, se observa que los *clusters* de puesto son también dominados por estos grupos. Finalmente, podemos indicar que los sectores de actividad de los trabajadores cualificados, y de los puestos que éstos ocupan, se concentran fundamentalmente en las actividades sanitarias y servicios sociales, educación, Administración Pública y otras actividades empresariales.

A continuación, nuestro análisis muestra algunos ejemplos de *biclusters* (véase el cuadro 5) que, aunque no dan lugar a un volumen de colocaciones demasiado

<i>Provincia puesto</i>	<i>Provincia puesto (%)</i>	<i>Municipio trabajador</i>	<i>Municipio trabajador (%)</i>	<i>Municipio puesto</i>	<i>Municipio puesto (%)</i>
Madrid	43.8	Municipio ≤ 40 000 A Coruña	50.1	Ferrol	29.3
A Coruña	29.3	Municipio ≤ 40 000 Pontevedra	25.7	Vigo	26.8
Pontevedra	26.8	Ferrol	10.1	Alcobendas	21.9
		Vigo	7.2	Madrid	21.9
		Municipio ≤ 40 000 Lugo	3.4		
		Otros	3.6		
Barcelona	40.3	Las Palmas	51.6	Barcelona	25.8
Madrid	22.6	Municipio ≤ 40 000 Santa Cruz de Tenerife	22.6	Madrid	21.0
Córdoba	19.4	Municipio ≤ 40 000 Las Palmas	9.7	Córdoba	19.4
Granada	8.1	Arona	6.5	Badalona	11.3
Ávila	6.5	Telde	4.8	Granada	8.1
Otros	3.2	Otros	4.8	Otros	14.5
Pontevedra	40.5	Municipio ≤ 40 000 Pontevedra	17.3	Vigo	22.1
A Coruña	39.9	Vigo	17.3	Municipio ≤ 40 000 A Coruña	17.3
Ourense	12.2	Municipio ≤ 40 000 A Coruña	17.0	A Coruña	16.1
Lugo	7.4	A Coruña	10.8	Municipio ≤ 40 000 Pontevedra	13.6
		Municipio ≤ 40 000 Ourense	8.2	Ourense	6.5
		Otros	29.5	Otros	24.4
Bizkaia	54.1	Municipio ≤ 40 000 Bizkaia	21.2	Municipio ≤ 40 000 Bizkaia	23.2
Álava	23.5	Vitoria	18.8	Bilbao	23.2
Gipuzkoa	12.9	Bilbao	14.4	Vitoria	19.7
Navarra	5.0	Municipio ≤ 40000 Gipuzkoa	11.5	Municipio ≤ 40 000 Gipuzkoa	9.1
La Rioja	4.4	Getxo	6.2	Municipio ≤ 40 000 Álava	3.8
		Otros	27.9	Otros	20.9

Cuadro 5, continuación...

Cluster de segmentos de trabajador	Cluster de segmentos de puesto	Ocupación trabajador	Ocupación trabajador (%)	Ocupación puesto	Ocupación puesto (%)
		Oficiales 1ª y 2ª	45.0	Oficiales 1ª y 2ª	52.1
		Oficiales 3ª y especialistas	31.5	Oficiales 3ª y especialistas	32.7
		Peones y asimilados	19.9	Peones y asimilados	15.2
99	35	Auxiliares administrativos	1.1		
		Subalternos	1.1		
		Otros	1.3		
		Ingenieros, licenciados, alta dirección	100.0	Ingenieros, licenciados, alta dirección	100.0
96	140				
		Ingenieros técnicos, ayudantes titulados	100.0	Auxiliares administrativos	19.8
				Oficiales 1ª y 2ª	18.1
112	33			Oficiales administrativos	17.6
				Oficiales 3ª y especialistas	13.6
				Ayudantes no titulados	9.6
				Otros	21.2
		Ingenieros técnicos, ayudantes titulados	60.0	Auxiliares administrativos	29.4
		Ingenieros, licenciados, alta dirección	38.8	Oficiales administrativos	20.3
9	4	Jefes administrativos y de taller	0.6	Oficiales 3ª y especialistas	16.5
		Ayudantes no titulados	0.6	Oficiales 1ª y 2ª	13.2
				Peones y asimilados	12.4
				Otros	8.2

elevado, se caracterizan por la existencia de cierto desajuste o *mismatch*²² en el emparejamiento, ya sea espacial u ocupacional, es decir, los trabajadores (al comienzo de su búsqueda) y los puestos se sitúan, hasta cierto punto, en zonas geográficas o en ocupaciones diferentes, por lo que los trabajadores requieren de una movilidad geográfica u ocupacional. Así, los dos primeros *bichusters* mues-

22 El *mismatch* es un concepto empírico que mide el grado de desajuste oferta-demanda en el mercado de trabajo a través de dos atributos fundamentales, el ocupacional y el espacial; ambos desajustes se pueden resolver a través de la movilidad de los trabajadores (o de las empresas) en una dirección u otra. Un trabajo interesante que analiza el *mismatch* existente dentro de un *cluster* laboral es el de Currid y Stolarick (2010). A un nivel más teórico, véanse los trabajos de Petrongolo y Pissarides (2001) y Shimer (2007).

<i>Actividad trabajador</i>	<i>Actividad trabajador (%)</i>	<i>Actividad puesto</i>	<i>Actividad puesto (%)</i>
Transporte marítimo	55.3	Transporte marítimo	58.6
Otras actividades empresariales	8.3	Comercio minorista	7.4
Fabricación de otro material de transporte	6.0	Otras actividades empresariales	6.5
Comercio minorista	5.1	Fabricación de otro material de transporte	6.5
Construcción	4.0	Industria de la construcción de maquinaria y equipo mecánico	3.6
Otros	21.3	Otros	17.4
Actividades sanitarias y servicios sociales	90.3	Actividades sanitarias y servicios sociales	100.0
Administración Pública	4.8		
Comercio mayorista	1.6		
Educación	1.6		
Comercio minorista	1.6		
Educación	28.9	Otras actividades empresariales	22.7
Actividades sanitarias y servicios sociales	16.7	Educación	15.6
Otras actividades empresariales	13.9	Hostelería	10.5
Administración Pública	11.9	Comercio minorista	9.9
Actividades asociativas	6.2	Administración Pública	9.3
Otros	22.4	Otros	32.0
Educación	33.8	Otras actividades empresariales	22.6
Otras actividades empresariales	13.5	Hostelería	14.7
Actividades sanitarias y servicios sociales	7.4	Educación	11.2
Administración Pública	6.8	Comercio minorista	5.9
Comercio mayorista	4.7	Actividades informáticas	4.7
Otros	33.8	Otros	40.9

tran un desajuste espacial. En el primer *bicluster* se observa un movimiento de trabajadores desde provincias gallegas hacia puestos situados en Madrid, y en el segundo *bicluster* la movilidad va desde las provincias de Las Palmas de Gran Canaria y Tenerife hacia provincias como Madrid, Barcelona y Córdoba. Se trata de mercados donde el empleo masculino tiene más peso y donde los trabajadores tienen en su mayoría 30 años o más. El primer *bicluster* corresponde principalmente a trabajadores de cualificación media, siendo el sector de actividad principal el de transporte marítimo. El segundo *bicluster* corresponde a trabajadores con estudios superiores que se dedican fundamentalmente a actividades sanitarias y de servicios sociales.

El tercer y el cuarto *bicluster* constituyen ejemplos de *mismatch* ocupacional: los trabajadores pertenecen sobre todo a los dos grupos de más elevada cualificación, los cuales no aparecen por el lado de los puestos entre las cinco categorías de mayor representación; este hecho constituye una muestra de “sobrecualificación” en el empleo —aunque hay que tener en cuenta que la propensión al emparejamiento de los correspondientes *clusters* de trabajador y de puesto es relativamente reducida—. El tercer *bicluster* se sitúa principalmente en Galicia, mientras que el cuarto lo hace en el País Vasco. Se trata de mercados donde el empleo femenino tiene un mayor peso y donde los trabajadores tienen en su mayoría entre 30 y 44 años. Los sectores de actividad con mayor representación en estos dos *biclusters* son educación, otras actividades empresariales, Administración Pública y actividades sanitarias y servicios sociales.

UN ANÁLISIS DE LAS DURACIONES DEL EMPLEO POR BICLUSTERS

El análisis desarrollado en la sección anterior se puede completar con una descripción de la duración de los episodios de empleo observados en los diferentes *biclusters*. El cuadro 6 ofrece información sintetizada de dichas duraciones tanto en los *biclusters* descritos en los cuadros anteriores (3, 4 y 5) como en seis nuevos *biclusters*, de los cuales tres se caracterizan por tener el mayor número de colocaciones de duración igual o superior a dos años, y los otros tres se caracterizan por tener el mayor porcentaje de colocaciones de duración igual o superior a dos años de entre todos aquellos *biclusters* que han generado durante el periodo considerado más de 1 000 colocaciones —garantizamos de esta forma el estudio de *biclusters* que generan un número significativo de colocaciones—. Las duraciones del empleo han sido divididas en seis intervalos: duraciones inferiores a un mes; superiores o similares a un mes e inferiores a tres meses; duraciones entre tres meses y menos de seis meses; entre seis meses y menos de un año; entre un año y menos de dos años; y duraciones iguales o superiores a dos años. Hay que tener en cuenta que la muestra empleada se centra en el flujo de altas laborales producidas desde el 1 de enero de 2011 hasta el 31 de diciembre de 2013, por lo que puede haber episodios de empleo que no han concluido todavía cuando se alcanza dicho momento final, existiendo, por tanto, censura por la derecha en nuestros datos.

Cuadro 6
Bicluster del mercado laboral español.
Duración del empleo

	Cluster de segmentos de trabajador	Cluster de segmentos de puesto	Propensión emparejamiento	Colocaciones bicluster	Duración < 1 mes (%)	1 mes ≤ duración < 3 meses (%)	3 meses ≤ duración < 6 meses (%)	6 meses ≤ duración < 1 año (%)	1 año ≤ duración < 2 años (%)	Duración ≥ 2 años (%)
<i>Biclusters del cuadro 3.</i> Elevadas frecuencias de trabajador y de puesto y alta propensión al emparejamiento	75	164	7.4	136 028	58.1	14.5	9.7	9.8	5.1	2.9
	41	73	7.6	118 349	58.8	13.8	10.2	9.9	4.7	2.6
	21	206	9.7	85 484	64.5	18.3	7.5	5.9	2.3	1.5
	36	117	15.0	54 040	55.1	19.2	11.7	9.6	3.0	1.4
	30	101	14.7	91 397	56.0	17.8	11.7	9.4	3.3	1.8
<i>Biclusters del cuadro 4.</i> Trabajadores cualificados	88	158	34.7	17 428	68.7	11.3	7.5	6.2	4.0	2.3
	54	90	60.1	9 857	70.0	10.1	9.8	5.4	2.5	2.3
	90	141	96.0	3 899	80.2	5.6	4.4	3.8	3.8	2.1
	9	18	109.3	3 732	71.2	11.4	5.8	5.3	3.5	2.7
	121	203	47.0	3 079	33.6	20.1	17.7	15.8	7.2	5.6
<i>Biclusters del cuadro 5.</i> Desajuste o <i>mismatch</i> en el emparejamiento	112	30	169.6	5 865	86.6	6.2	2.7	2.5	0.9	1.0
	99	35	16.4	447	22.8	33.1	22.6	13.6	4.7	3.1
	96	140	2.4	62	45.2	24.2	11.3	11.3	3.2	4.8
	112	33	1.0	353	43.1	17.6	13.9	13.9	6.5	5.1
<i>Biclusters con un elevado número de colocaciones de duración ≥ 2 años</i>	9	4	1.2	340	41.8	19.7	15.9	11.5	7.4	3.8
	41	74	5.7	70 969	47.8	12.8	11.3	12.2	9.0	7.0
	75	164	7.4	136 028	58.1	14.5	9.7	9.8	5.1	2.9
<i>Biclusters con más de 1 000 colocaciones y mayor porcentaje de duración ≥ 2 años</i>	41	73	7.6	118 349	58.8	13.8	10.2	9.9	4.7	2.6
	75	149	5.9	1 352	7.0	8.1	10.4	16.8	15.7	42.1
	9	2	46.1	1 784	23.6	11.2	11.9	18.8	17.8	16.6
	75	162	3.9	4 475	21.5	12.1	13.3	21.2	17.6	14.4
Total de colocaciones (%)					59.1	15.4	10.0	9.1	3.9	2.5

La distribución de las colocaciones por duraciones resulta en general decreciente, de manera que la mayor parte de los empleos muestran una duración inferior a un mes —en concreto, un 59.1% del total de la muestra—. Dicha preponderancia se observa claramente en dos *biclusters* de trabajadores cualificados, el 112-30, situado en Galicia, y el 90-141, situado mayoritariamente en Valencia, donde las colocaciones de escasa duración superan el 80%; ambos focos de generación de empleo se corresponden con el sector de actividad sanitario.

A pesar de la tendencia descrita respecto a la temporalidad del empleo, existen también ciertos *biclusters* donde las duraciones iguales o superiores a dos años son relativamente importantes. Esto sucede, por ejemplo, en los *biclusters* 41-74, 75-164 y 41-73, que pese a estar compuestos mayoritariamente por episodios de empleo de corta duración —los porcentajes de colocaciones de menos de un mes rondan o superan el 50% en los tres casos—, son los que presentan un mayor número absoluto de colocaciones con duración superior a dos años. El primero y el tercero se sitúan fundamentalmente en la provincia de Madrid, mientras que el segundo lo hace principalmente en Barcelona. El primer *bicluster* corresponde sobre todo a trabajadores de los grupos de auxiliares administrativos y de oficiales administrativos, mientras que en los otros dos resultan más importantes los grupos de peones y asimilados y de oficiales de 1ª y 2ª. En los tres mercados destacan sectores de actividad como el de otras actividades empresariales, el de las actividades de ocio y el de la hostelería.

Otros *biclusters* interesantes desde el punto de vista de la estabilidad del empleo son el 75-149, el 9-2 y el 75-162, que se caracterizan por presentar el mayor porcentaje de colocaciones con duración superior o igual a dos años entre los *biclusters* con más de 1 000 colocaciones —estos porcentajes son, respectivamente, 42.1, 16.6 y 14.4—. El primero y el tercero se sitúan fundamentalmente en la provincia de Barcelona y se corresponden con trabajadores de cualificación media (jefes administrativos y de taller, auxiliares administrativos y ayudantes no titulados); el primero está vinculado al sector de la intermediación financiera, y el tercero se vincula a los sectores de otras actividades empresariales, comercio mayorista y comercio minorista. Por su parte, el segundo de estos tres *biclusters* se ubica principalmente en el País Vasco y se corresponde con trabajadores titulados que encuentran empleo en los sectores de otras actividades empresariales y Administración Pública.

Como ya hemos señalado, un aspecto característico de la economía española es el elevado grado de temporalidad del empleo, sobre todo del flujo de generación de nuevos empleos. El fenómeno de la temporalidad guarda una estrecha relación con el hecho de que un número significativo de trabajadores experimentan una dinámica consistente en encadenar un número considerable de relaciones laborales de corta duración, ya sea en la misma o en diferentes empresas —sobre el fenómeno de la temporalidad y la recurrencia en el empleo en España véanse, por ejemplo, los trabajos de Dolado, García-Serrano y Jimeno (2002), Alba-Ramírez, Arranz y Muñoz-Bullón (2007) y Arranz y García-

Serrano (2014)—. Evidentemente, este rasgo idiosincrásico de nuestro mercado de trabajo debe tener su reflejo en el mapa de *biclusters* obtenido, de manera que aquellos *biclusters* con mayor temporalidad deberían mostrar un mayor porcentaje relativo de empleo recurrente. Para mostrar esta idea, hemos analizado el número de contratos acumulado por un trabajador (y su duración media) en el total de la muestra y en un par de *biclusters* que difieren claramente en su grado de temporalidad. En el total de la muestra se observa que el 52.2% de los trabajadores ha tenido sólo uno o dos contratos a lo largo del periodo analizado (contratos cuya duración media es igual a 312.7 días²³), mientras que el porcentaje de trabajadores con más de diez contratos es del 8.8% (contratos de duración media igual a 17.2 días). Sin embargo, estos valores son, respectivamente, del 48.7% (duración media 383.2 días) y del 32.3% (duración media 11.5 días) en el *bicluster* sanitario de elevada temporalidad 112-30, situado en Galicia, es decir, el porcentaje de trabajadores que han experimentado once o más contratos aumenta significativamente, respecto al total de la muestra, en este *bicluster*. Por su parte, los valores analizados son respectivamente 96.5% (duración media de 566.3 días) y 0.3% (duración media de 59.3 días) en el *bicluster* de empleo relativamente estable 75-149, situado en Barcelona y dedicado principalmente a la intermediación financiera; en este *bicluster*, la mayor parte de los trabajadores solamente ha tenido uno o dos contratos entre 2011 y 2013. Como se puede apreciar, existe una notable heterogeneidad a este respecto entre los *biclusters*.

Los casos que hemos seleccionado y comentado en la sección anterior y en esta sección resultan ilustrativos de la utilidad de nuestra metodología para conocer la estructura de los yacimientos de empleo y orientar el diseño de las políticas activas en este campo. Asimismo, hay que tener en cuenta que todo el análisis desarrollado en este artículo se podría ampliar, al menos, en dos direcciones. Por un lado, se podrían incluir nuevas variables, procedentes de la MCVL, que permitieran conocer mejor la estructura de los segmentos, de los *clusters* y de los mercados laborales (*biclusters*) identificados; variables como el tipo de contratación, el tipo de jornada, el nivel salarial promedio, etcétera. Por otro lado, resultaría posible abordar el análisis desde un punto de vista dinámico, ya que la MCVL contiene información fiable de altas laborales desde 2005. Por lo

23 Téngase en cuenta que únicamente un 17.8% de los episodios de empleo correspondientes a los trabajadores que han tenido sólo uno o dos contratos a lo largo del periodo muestra una duración inferior a un mes.

tanto, resultaría posible, por ejemplo, analizar el efecto del ciclo económico en los diferentes segmentos laborales, o en los *clusters* que éstos forman, e incluso tratar de identificar segmentos o *clusters* emergentes a partir de la evolución de sus colocaciones, conociendo, además, toda la dinámica de las propensiones al emparejamiento observadas.

Finalmente, nos gustaría destacar la versatilidad de nuestra metodología, ya que podría ser aplicada a mercados laborales de otras economías, o a otras bases de datos laborales, o incluso ser aplicada a bases de datos no laborales pero que estén basadas en emparejamientos —mercados financieros, inmobiliarios, etc.—.

CONCLUSIONES

El presente trabajo pretende mostrar cómo a partir de bases de datos individuales con millones de colocaciones es posible obtener una visión sintética de “quién se empareja con quién” en el mercado de trabajo, dando lugar a un mapa de *biclusters* laborales o mercados específicos generadores de empleo. Estos mercados están formados por grupos de segmentos de trabajador y de segmentos de puesto que tienden a emparejarse entre sí en mayor medida de lo que sucedería en un escenario donde los emparejamientos funcionaran de forma puramente aleatoria. La metodología de agrupamiento propuesta, ejemplificada mediante el uso de la MCVL (periodo 2011-2013), permite procesar la ingente cantidad de información existente sobre las colocaciones ya producidas, con el fin último de generar una “hoja de ruta” sobre los diferentes focos de generación de empleo del mercado de trabajo español. Esta forma novedosa de estructurar la información puede redundar en una reducción del *mismatch* laboral, tal y como es entendido en los modelos de emparejamiento.

Nuestro proceso de síntesis informativa puede estructurarse en tres etapas. En primer lugar, generamos segmentos (pequeños grupos) de trabajadores y de puestos atendiendo a las características de los trabajadores y de los puestos en cada colocación observada; esta segmentación resulta coherente con la forma en que los modelos de búsqueda (por los dos lados del mercado) y emparejamiento entienden el funcionamiento del mercado de trabajo. En concreto, nos centramos en cinco características del trabajador que pueden provocar segmentación en el emparejamiento: municipio, grupo de ocupación, tramo de edad, sexo y sector de actividad; y en tres características del puesto cubierto: municipio, grupo de ocupación y sector de actividad. Cada colocación dará lugar a

un segmento conjunto (de ocho características). Interpretamos que el mercado de trabajo está segmentado cuando los trabajadores de un segmento específico tienden a emparejarse de forma más o menos intensa con ciertos segmentos de puesto específicos en comparación con una asignación aleatoria. En relación con esta idea de segmentación, proponemos dos medidas empíricas: la “propensión al emparejamiento” entre un segmento de trabajador y un segmento de puesto, y el “grado de similitud” entre dos segmentos de trabajador (o dos segmentos de puesto). En segundo lugar, hemos procedido a realizar una agrupación de los segmentos de trabajador, por un lado, y de los de puesto por otro, atendiendo fundamentalmente a la “similitud” de sus respectivos municipios; de esta forma pretendemos conseguir una matriz de emparejamientos entre segmentos con un mayor número de colocaciones o frecuencias en cada cruce de segmento de trabajador y de puesto —tratamos de resolver un problema de excesivas frecuencias muy bajas o nulas—. Aumentando las frecuencias conjuntas observadas, mejoramos nuestro análisis, ya que aumenta la fiabilidad de las medidas de propensión al emparejamiento y de similitud entre segmentos. Finalmente, los 9 170 segmentos de trabajador y los 9 208 segmentos de puesto obtenidos en la agrupación anterior han sido respectivamente agrupados en 256 *clusters* jerárquicos en función de las similitudes de los diferentes segmentos entre sí, *clusters* que son posteriormente combinados según sus emparejamientos para generar *biclusters* (*clusters* de trabajador con una cierta propensión a emparejarse con *clusters* de puesto) que permitan obtener una visión sintética de la segmentación existente en el mercado de trabajo. Hay que tener en cuenta que las sucesivas agrupaciones de segmentos que llevamos a cabo en este trabajo no suponen pérdida alguna de información, es decir, conocemos en todo momento los segmentos iniciales u originales que hay dentro de cada agrupación y las frecuencias o colocaciones de éstos; de hecho, el análisis descriptivo de cada *cluster* laboral se basa en dichos segmentos de partida.

Los resultados obtenidos, aparte de confirmar la elevada segmentación y excesiva temporalidad existente en el mercado de trabajo español, indican que los principales mercados laborales específicos detectados muestran un grado de desajuste o *mismatch* relativamente reducido tanto a nivel geográfico como ocupacional. Así, los trabajadores y los puestos de aquellos *biclusters* generadores de empleo, y con una alta propensión al emparejamiento, se sitúan aproximadamente en la misma área geográfica —si bien los puestos se concentran algo más en las capitales de provincia que los trabajadores, lo que podría apuntar

hacia cierto grado de *commuting*— y en grupos de ocupación similares —aunque parece haber algún indicio de “sobrecualificación” en el empleo, no parece ser un problema de peso en este tipo de *biclusters*—. Sin embargo, este escenario de ajuste parece no cumplirse en ciertos mercados donde las propensiones al emparejamiento entre *clusters* son algo más moderadas; así, en estos *biclusters* la movilidad geográfica u ocupacional del trabajador podría jugar un papel más relevante.

Nuestra aproximación puede ser muy versátil en su aplicación, ya que se puede orientar hacia el problema específico que se desee estudiar; por ejemplo, permite analizar de forma sintetizada el problema de la falta de estabilidad en el empleo en los diferentes ámbitos laborales. La división, de forma endógena (según los datos manejados, en nuestro caso la MCVL), del mercado de trabajo global en un conjunto de mercados específicos, generadores de empleo, posibilita el diseño de políticas laborales más desagregadas y ajustadas, por ejemplo, de movilidad geográfica y ocupacional; políticas que van a depender de la estructura laboral que encontremos dentro de cada mercado o *bicluster*. Evidentemente, la calidad del diseño de estas políticas dependerá del grado de información que se tenga sobre cada mercado específico detectado. Nuestro trabajo ha intentado mostrar la metodología propuesta de detección de esos mercados laborales, así como varias aplicaciones representativas.

REFERENCIAS

- Agarwal, N. (2015). An empirical model of the medical match. *American Economic Review*, 105(7), pp. 1939-1978.
- Alba-Ramírez, A., Arranz J.M. y Muñoz-Bullón, F. (2007). Exits from unemployment: Recall or new job. *Labour Economics*, 14(5), pp. 788-810.
- Álvarez de Toledo, P., Núñez, F. y Usabiaga, C. (2008). La función de emparejamiento y el mercado de trabajo español. *Revista de Economía Aplicada*, 16(48), pp. 5-35.
- Álvarez de Toledo, P., Núñez, F. y Usabiaga, C. (2011). An empirical analysis of the matching process in Andalusian public employment agencies. *Hacienda Pública Española*, 198(3), pp. 67-102.
- Álvarez de Toledo, P., Núñez, F. y Usabiaga, C. (2013). Análisis “cluster” de los flujos laborales andaluces. *Revista de Estudios Regionales*, 97, pp. 195-221.

- Álvarez de Toledo, P., Núñez, F. y Usabiaga, C. (2014). An empirical approach on labour segmentation. Applications with individual duration data. *Economic Modelling*, 36, pp. 252-267.
- Arranz, J.M. y García-Serrano, C. (2011). Are the mcvl tax data useful? Ideas for mining. *Hacienda Pública Española*, 199(4), pp. 151-186.
- Arranz, J.M. y García-Serrano, C. (2014). Duration and recurrence of unemployment benefits. *Journal of Labor Research*, 35(3), pp. 271-295.
- Barnichon, R. y Figura, A. (2015). Labor market heterogeneity and the aggregate matching function. *American Economic Journal: Macroeconomics*, 7(4), pp. 222-239.
- Becker, G.S. (1973). A theory of marriage: Part I. *Journal of Political Economy*, 81(4), pp. 813-846.
- Boyd, D., Lankford, H., Loeb, S. y Wyckoff, J. (2013). Analyzing the determinants of the matching of public school teachers to jobs: Disentangling the preferences of teachers and employers. *Journal of Labor Economics*, 31(1), pp. 83-117.
- Cebrián, I. y Toharia, L. (2008). La entrada en el mercado de trabajo. Un análisis basado en la mcvl. *Revista de Economía Aplicada*, 16(número extraordinario 1), pp. 137-172.
- Cebrián, I., Moreno, G. y Toharia, L. (2011). La estabilidad laboral y los programas de fomento de la contratación indefinida. *Hacienda Pública Española*, 198(3), pp. 103-127.
- Cotterman, R. y Peracchi, F. (1992). Classification and aggregation: An application to industrial classification in cps data. *Journal of Applied Econometrics*, 7(1), pp. 31-51.
- Currid, E. y Stolarick, K. (2010). The occupation-industry mismatch: New trajectories for regional cluster analysis and implications for economic development. *Urban Studies*, 47(5), pp. 337-362.
- Dolado, J.J., García-Serrano, C. y Jimeno, J.F. (2002). Drawing lessons from the boom of temporary jobs in Spain. *The Economic Journal*, 112(480), pp. 270-295.
- Everitt, B.S., Landau, S., Leese, M. y Stahl, D. (2011). *Cluster Analysis*. Quinta edición. Nueva York: Wiley.
- Fox, J.T. (2008). *Estimating matching games with transfers* (Working Paper no. 14382). Disponible a través de National Bureau of Economic Research (NBER) <<http://www.nber.org/papers/w14382.pdf>>.
- Fox, J.T. (2010). Identification in matching games. *Quantitative Economics*, 1(2), pp. 203-254.

- Gale, D. y Shapley, L.S. (1962). College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1), pp. 9-14.
- García-Cintado, A., Romero-Ávila, D. y Usabiaga, C. (2014). *Spanish Regional Unemployment. Disentangling the Sources of Hysteresis*. Berlín: Springer (Springer-Briefs in Economics).
- Jain, A.K., Murty, M.N. y Flynn, P.J. (1999). Data clustering: A review. *ACM Computing Surveys*, 31(3), pp. 265-323.
- Kaufman, L. y Rousseeuw, P.J. (2005). *Finding Groups in Data: An Introduction to Cluster Analysis*. Segunda edición. Nueva York: Wiley.
- Lapuerta, I. (2010). *Claves para el trabajo con la Muestra Continua de Vidas Laborales* (DemoSoc Working Paper no. 2010-37). Universitat Pompeu Fabra, Barcelona.
- Liu, B. (2011). *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*. Berlín: Springer-Verlag.
- Núñez, 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*. Sevilla: Centro de Estudios Andaluces.
- Panteli, A., Boutsinas, B. y Giannikos, I. (2014). On set covering based on biclustering. *International Journal of Information Technology & Decision Making*, 13(5), pp. 1029-1049.
- Petrongolo, B. y Pissarides, C.A. (2001). Looking into the black box: A survey of the matching function. *Journal of Economic Literature*, 39(2), pp. 390-431.
- Pissarides, C.A. (2000). *Equilibrium Unemployment Theory*. Segunda edición. Cambridge, MA: The MIT Press.
- Pissarides, C.A. (2011). Equilibrium in the labor market with search frictions. *American Economic Review*, 101(4), pp. 1092-1105.
- Shimer, R. (2007). Mismatch. *American Economic Review*, 97(4), pp. 1074-1101.