Labour Market Segmentation, Clusters, Mobility and Unemployment Duration with Individual Microdata

Álvarez de Toledo, Pablo and Núñez, Fernando and Usabiaga, Carlos

Seville University, Seville University, Pablo de Olavide University, Seville

January 2013

LABOUR MARKET SEGMENTATION, CLUSTERS, MOBILITY AND UNEMPLOYMENT DURATION WITH INDIVIDUAL MICRODATA

Pablo Álvarez de Toledo. **Universidad de Sevilla**.
Fernando Núñez. **Universidad de Sevilla**.
Carlos Usabiaga. **Universidad Pablo de Olavide**.

Abstract

This article proposes empirical tools to account for the role of heterogeneities in the labour matching process, and shows an application to the Andalusian labour market which relies on individual microdata. Firstly, by considering that the labour market is segmented when workers of a specific group have greater probability of matching with specific job groups, we propose two empirical measures related to this idea: propensity to match, and segmentation in worker and job groups. Secondly, we use a clustering methodology, based on a similarity measure, to obtain a better overview of the structure of the labour market. Thirdly, we propose a measure of mobility based on our similarity measure, and estimate a regression model that relates mobility to worker and job characteristics and to the economic cycle. Finally, these tools are included in an unemployment duration model. The proposed methodology may be useful in labour intermediation by helping seekers to follow a 'roadmap' of successful paths.

Keywords: Heterogeneity, Local labour markets, Cluster analysis, Mobility, Unemployment duration.

JEL codes: J42, J61, J62, J64, C38.

ACKNOWLEDGMENTS

We are grateful to the Andalusian Public Employment Agency (Servicio Andaluz de Empleo, SAE) for the provision of the data. We thank the Centro de Estudios Andaluces for the funding received specifically for this line of research (Project PRY010/10). C. Usabiaga also acknowledges additional research funding from the Andalusian Government (Excellence Project SEJ-4546, PAI Group SEJ-513), the Spanish Ministry of Science and Innovation (ECO2009-13357), and the Spanish Ministry of Economics and Competitiveness (ECO2012-35430).

E-mail de contacto (Carlos Usabiaga): cusaiba@upo.es

1. Introduction

In the labour market, workers seeking jobs and vacant jobs offered by employers are heterogeneous in aspects as skills, geographical location, gender, age, payment, working time, attitude, taste, and many others. These heterogeneities lead to the concept of mismatch: "Mismatch is an empirical concept that measures the degree of heterogeneity in the labour market across a number of dimensions, usually restricted to skills, industrial sector, and location. Large differences in the skills possessed by workers and those required by firms would lengthen the time that it takes to match a given group of workers to a given group of firms, as agents search for a good match among the heterogeneous group. Industrial sector matters in matching because of industry-specific skills that may not be picked up by generally available measures of skills. Finally, location influences matching because of imperfect labour mobility." (Petrongolo and Pissarides 2001, 399-400).

In this paper, we propose some empirical tools to account for the role of heterogeneities in the labour matching process, and we then make use of them in an application to the Andalusian labour market[1], which relies on a database of individual microdata of considerable size. We begin by dividing the workers, the jobs and the (worker-job) matches into highly detailed groups according to their characteristics (location and skills in our application). Ideally, the detailed division should allow us to consider the groups obtained as homogeneous or almost homogeneous, and the large size of the database should enable data in each group to be sufficiently numerous as to be statistically representative.

The nature of our data, with information on vacancies, unemployed workers and job placements, links up our work directly with the theoretical concept of matching function. This function is intended to represent heterogeneities, frictions, and information imperfections and to capture the implications of the costly trading process without the need to make the heterogeneities and other features that give rise to it explicit. Instead of representing frictions more specifically according to their origin and their type, we lump them all together into an aggregate function. Therefore, the matching function does not assume that workers and jobs are homogeneous[2]; it simply omits to make the heterogeneities explicit. Without heterogeneities (zero mismatch), the matching function would not exist and jobs and workers would match instantaneously (Pissarides 2000, 3-4, 22, Pissarides 2008, Shimer 2007, 1077, Petrongolo and Pissarides 2001, 400)[3].

Considerable work has been carried out in an effort to open the 'black box' of the matching process and to render the heterogeneities inside the matching function explicit. Island, urn-ball, taxicab, queuing, stock-flow (or marketplace) and mismatch models, have all explored different types of frictions, extending the search theory of the labour market to allow for worker and firm

[1] This application to Andalusia is also interesting because it is the most populated Spanish region and persistently one of the European regions with the highest unemployment rate. In this region the main problems of the Spanish labour market – Bentolila et al. (2012) – are exacerbated.

[2] Several authors seem to state this. For example Yashiv 2007, 1872: "In the basic model all workers and jobs are assumed homogeneous …" and Brown et al. 2009, 4: "In many conventional search models that use a matching function, workers and jobs are treated as if each group were homogeneous and randomly matched".

[3] There exists an extensive literature that surveys search and matching theories applied to labour economics and the matching function; see, for example, Devine and Kiefer (1991), Mortensen and Pissarides (1999), Pissarides (2000), Petrongolo and Pissarides (2001), Rogerson et al. (2005), and Yashiv (2007).

heterogeneity and for micro-foundations of the matching process[4]. As a rule, the labour market, or workers and jobs, are divided into parts (local labour markets, locations, islands, queues, worker-job pairs acceptable or unacceptable to match productively, stock (old)-flow (new) workers and jobs), which are then treated as if each part were homogeneous.

Our work is not meant to extend or evaluate theoretical models of labour matching, but instead it tries to handle empirically important elements involved in these models – heterogeneities and segmentation –. We start out our analysis by considering that the labour market is segmented when the workers of a particular group have greater probability of matching with certain job groups than with others. Otherwise, we denominate it as a non-segmented or purely random (PR) labour market. Obviously, the heterogeneities of workers and jobs are the reason that the labour market is segmented, since with completely homogeneous workers and jobs it would be a PR labour market, but the two concepts – heterogeneities and segmentation – are distinct[5]. We propose a measure of the degree of segmentation of each group and another measure of the propensity to match between workers and jobs depending on the groups to which they belong. As might be expected, our data show a very high degree of segmentation for the vast majority of groups.

Since highly detailed division results in a very large number of groups, which may be difficult to interpret, we use a clustering[6] methodology, based on a similarity measure, to obtain a better overview of the structure of the labour market and to obtain a smaller number of clusters ('groupings of groups'). Cluster analysis enables, as far as possible, subjective or 'a priori' grouping criteria to be avoided: in our case, this would be the case, for example, if, for locations, municipalities were grouped in provinces and regions, or if, for skills, classifications with fewer digits for occupations or sectors of economic activity were used. Instead, we look for a measure of similarity adapted, in the most objective possible way, to the purpose of our research. In the context of the search and matching theories applied to labour economics, we consider that worker (job) groups are more similar the more they resemble in the way they match with job (worker) groups. Using this concept of similarity, we will show in which way the worker-job clusters with high propensity to match that are formed may be considered as labour market clusters. We present results obtained by applying this methodology to our data[7].

Mobility and unemployment duration are essential concepts in the search models that make the heterogeneities explicit by dividing the labour market into parts and specifying how workers (and jobs) move from one to another part[8]. We propose an empirical measure of mobility directly

[4] See, for example, Phelps (1970), Lucas and Prescott (1974) and Mortensen (2009) about island models, Petrongolo and Pissarides (2001) about urn-ball models, Lagos (2000) on the taxicab model, Gautier (2002) and Sattinger (2010) on queuing, Coles and Smith (1998) and Ebrahimy and Shimer (2010) about stock-flow models, and Shimer (2007) on mismatch models.

[5] In labour economics, the concept of market segmentation has also been used in a more restrictive sense than ours. This applies to the theory of dual labour markets – see for example Reich et al. (1973) – or to the branch of endogenous segmentation – Moreno-Galbis (2009) –.

[6] About cluster analysis see, for example, Cotterman and Peracchi (1992), who propose an application to an industrial classification, and the survey of Jain et al. (1999).

[7] More detailed results on local labour markets in Andalusia can be found in Álvarez de Toledo et al. (2012).

[8] Mobility has been studied from different perspectives in empirical literature: 'job' mobility – Topel and Ward (1992), Shimer (2007) and Barnichon and Figura (2011) –, 'interregional' mobility – Greenwood (1985), Pissarides and Wadsworth (1989), Faini et al. (1997), Ahn et al. (1999) and Arellano and Bover (2002) –, 'occupational' mobility – Miller (1984), McCall (1990), Neal (1995), Kambourov and Manovskii (2009) and Lalé (2012) –, and 'sectoral' mobility – Lilien (1982), Abraham and Katz (1986), Jovanovic and Moffitt (1990), Iglesias-Fernández and Llorente-Heras (2007) and Bachmann and Burda (2010) –.

related to our similarity measure, and then we estimate a multiple regression model that relates mobility in each worker-job match primarily to worker characteristics, and also to job characteristics and macroeconomic conditions. We use the results of the regression to estimate the 'a priori' workers' willingness to move. Our analysis ends up showing that the new empirical framework developed in this work can enhance the estimation of unemployment duration models in this field[9].

The rest of the paper is organised as follows. Section 2 analyses the concept of labour market segmentation and proposes some related empirical measures: propensity to match and segmentation in worker and job groups. Section 3 develops the clustering methodology and shows the structure of the labour market obtained by applying this methodology. Section 4 proposes a measure of mobility and estimates a regression model that relates this measure to worker and job characteristics and to macroeconomic conditions. The results are used to estimate the willingness of workers to move. Section 5 estimates an unemployment duration model making use of the tools obtained in the previous sections. Finally, Section 6 draws conclusions and suggests a number of possible applications of our methodology to active labour market policies.

2. Labour market segmentation

At any period t in time, each worker seeking for a job is assigned to one of the *n worker groups* W_{it} (i=1,2, … n), each vacant job is assigned to one of the *m job groups* J_{jt} (j=1,2, … m), and each of the matches formed with both, worker and job, is assigned to one of the *n x m joint groups* S_{ijt}. Each group is defined by the corresponding set of characteristics $\bar{W}_i, \bar{J}_j, \bar{S}_{ij}$ plus the t period ($\bar{S}_{ij}$ includes the characteristics of the worker $\bar{W}_i$ and the characteristics of the job $\bar{J}_j$ that are matched)[10]. In period t, the number of matches in each joint group, M_{ijt}, shows 'who matches with whom'[11]. The total number of matches, $M_t = \sum_{i=1}^{n}\sum_{j=1}^{m} M_{ijt}$, is the sum of matches for all the joint groups, and the number of matches in each worker and job group are, respectively, $M_{it} = \sum_{j=1}^{m} M_{ijt}$, $M_{jt} = \sum_{i=1}^{n} M_{ijt}$.

Although not necessarily so, it may be desirable in some cases to consider the same set of characteristics for both workers and jobs. In this case, we will have symmetric characteristics. Each worker group has a mirror job group with identical values for the set of characteristics and vice versa. In this case, we say that the joint group that corresponds to mirror worker and job groups is a mirror joint group. We also consider that two joint groups are symmetric if the worker group of one of them is the mirror of the job group of the other, and vice versa.

In period t, the sampling probabilities that a match occurs in the ij joint group, in the i worker group, and in the j job group, are, respectively

[9] There exists an extensive literature on unemployment duration and hazard rates. We can highlight the works of Lancaster (1979), Devine and Kiefer (1991), van den Berg and van Ours (1996), Machin and Manning (1999), Yashiv (2007), Shimer (2007, 2008), Kuo and Smith (2009) and Rogerson and Shimer (2011). For the Spanish economy, see for instance Alba-Ramírez (1994), Ahn et al. (1999), Carrasco (1999), Bover et al. (2002), Güell and Hu (2006), Güell and Petrongolo (2007), Carrasco and García-Pérez (2008) and Álvarez de Toledo et al. (2011).

[10] Barnichon and Figura (2011) formally use a similar type of labour market segmentation.

[11] Similarly, in two-sided matching games, a match production function governs who matches with whom. See, for example, Fox (2008).

$$p_{ijt} = M_{ijt} / M_t\,,\quad p_{it} = \sum_{j=1}^{m} p_{ijt}\,,\quad p_{jt} = \sum_{i=1}^{n} p_{ijt} \tag{1}$$

We consider that the labour market is segmented if the workers of a particular group have greater probability of matching with certain job groups than with others. If this is not the case, and the distributions of matches by worker and job groups are independent, then we consider that the labour market is non-segmented or a purely random (PR) labour market. In this case, the random joint estimated probability that a match occurs in the *ij* joint worker-job group is

$$\widehat{p_{ijt}} = p_{it}\, p_{jt} \tag{2}$$

which, in segmented labour markets, will be different, in general, from p_{ijt}. In the other extreme, if each worker group matches with only one single job group, we would have 'pure island' joint groups.

We can measure the particular *propensity to match* pm_{ijt} between the *i* worker group and the *j* job group in period *t*, as

$$pm_{ijt} = p_{ijt} / \widehat{p_{ijt}} \tag{3}$$

whereby pm_{ijt} is one[12] in PR labour markets. In segmented labour markets, it is greater than one when workers of the *i* worker group match with the jobs in the *j* job group 'over PR', and vice versa. For each period, by using $\widehat{p_{ijt}}$ as weights, the weighted sample mean of pm_{ijt} is one, and its weighted sample variance (zero in PR labour markets) is a measure of segmentation.

With symmetric characteristics, we can expect a high propensity to match in the mirror joint groups. It seems very likely that the workers of a certain group A (located in A, with skills A, etc.) have a high propensity to match with jobs of group A (located in A, requiring skills A, etc.). We can also expect a high positive correlation of the propensities to match of symmetric joint groups. If the workers of a certain group A have a high propensity to match with jobs of a certain group B, it seems likely that the workers of the group B have a high propensity to match with jobs of the group A.

We can measure the degree of segmentation for each of the worker groups in period *t* as follows. With non-segmentation, workers of the group W_{it} match with each of the J_{jt} job groups with random probabilities p_{jt} (equal to $\widehat{p_{ijt}} / p_{it}$), independent of *i*. However, due to segmentation, the effective sampling probabilities p_{ijt} / p_{it} (equal to $pm_{ijt}\, p_{jt}$) are dependent on *i*; the propensity to match pm_{ijt} being the ratio between these effective and random probabilities. We place the job groups in increasing order of pm_{ijt} for this particular *i*. If we represent the accumulated value of effective probabilities $pm_{ijt}\, p_{jt}$ against the accumulated value of random probabilities p_{jt}, then we obtain the slope-increasing solid line in Fig. 1[13]. With non-segmentation, we obtain the constant unitary slope dashed line. The more 'selective' the workers are, concentrating their matches on certain job groups, the more separated the two lines become. The proposed 'Gini type' measure of segmentation sg_{it} in the worker group W_{it} is the ratio of the areas A and A+B (=1/2) of the figure

$$sg_{it} = A/(A+B) = 2A = 1\text{-}2B \tag{4}$$

[12] Asymptotically, with an infinite sample size.

[13] Obviously, the sum over all the job groups, both for random and effective probabilities, must be one.

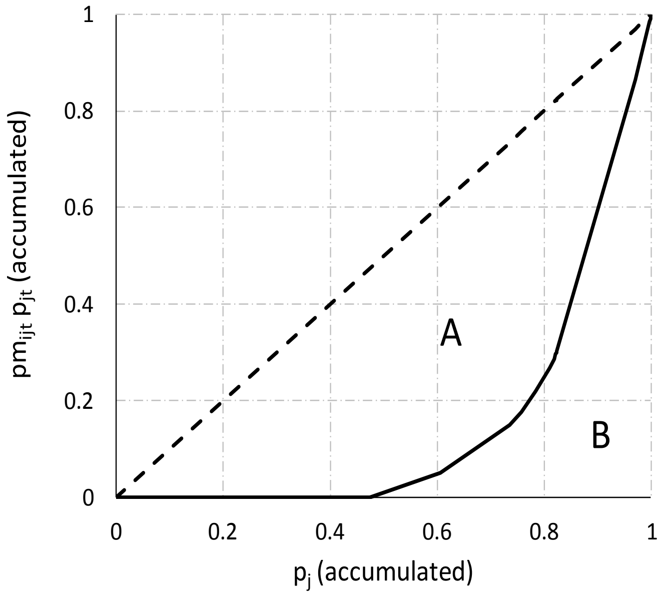


Fig. 1. Measure of segmentation in the worker group W_{it}.

With non-segmentation, A=0, B=1/2 and sg_{it}=0. With extreme segmentation, A approaches 1/2, B approaches zero, and sg_{it} approaches one. We can measure the degree of segmentation sg_{jt} for each of the job groups in period *t* in an analogous way.

Our data, and the empirical application that we have implemented, refer to the matches registered in the Andalusian Public Employment Agency (Servicio Andaluz de Empleo, SAE) in the four years 2007 to 2010[14]. The available information allows us to make a detailed division into groups, with the combination of various characteristics, and yet have enough data in each group to be statistically representative. Both for workers and for jobs, we have considered symmetric characteristics: location, defined by municipality (770 different municipalities); and skills, defined by occupation (787 different occupations), plus sector of economic activity (56 different sectors)[15]. During the four years analysed, there was a flow of more than 16 million registered matches, but all the values of the full set of characteristics are known in only just over 9 million matches, distributed between 2,848,974 different joint groups, 456,109 different worker groups and 261,167 different job groups, of which 119,614 are common groups that are part of both worker and job groups. Additionally, in the same period, our data include monthly registers of stocks of seeking workers and vacant jobs, with mean values over the whole period of 1,163,433 registered seeking workers and 18,542 registered vacant jobs. Many of the registered matches include workers and, especially, jobs not previously registered, but we know the characteristics of these jobs and workers by the match register. Finally, in order to manage some of our subsequent calculations in the cluster analysis, we have been forced to reduce the large amount of information available by selecting a sample of 1,587 common groups that appear in the 10,000 joint groups with the most matches. For these 1,587 groups, there are 1,906,828 matches distributed between 69,954 different joint groups.

[14] The availability of reliable individual data starts in 2007 – SISPE methodology –.

[15] For workers, location is usually their place of residence, and for jobs, where the work takes place. The skills are those possessed by the worker or required by the job.

In Fig. 2, we show the distribution of pm_{ij} with our data, considering a single four-year period. The figure shows that the labour market is clearly segmented, with almost all worker-job group combinations with zero propensity to match[16], and just a few with very high propensity. The weighted sample variance (49,899) also indicates a high degree of segmentation. As expected, in the mirror joint groups the propensity to match is much higher, with a weighted mean of 346.94 versus one for the total. Unsurprisingly, we also find a clear positive correlation (0.64) between the propensities to match of symmetric joint groups.

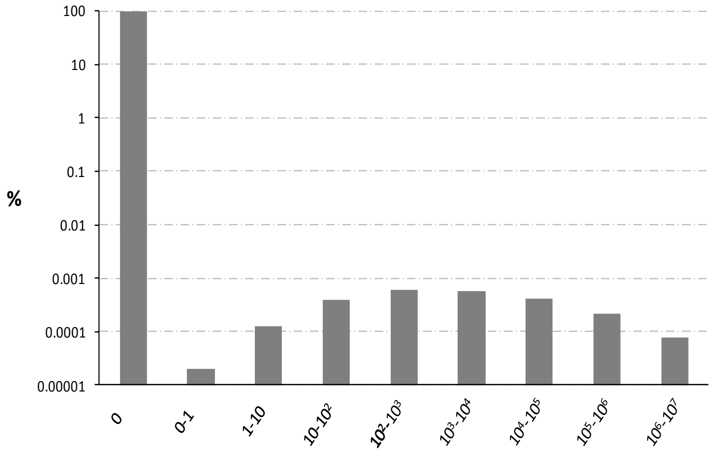


Fig. 2. Distribution of propensity to match.

In Fig. 3, we show the distribution of sg_i and sg_j, considering a single four-year period. Again, the figure shows that the labour market is clearly segmented, with the vast majority of the values of segmentation in worker and job groups very close to one (the mean value is 0.9982 for worker groups and 0.9995 for job groups).

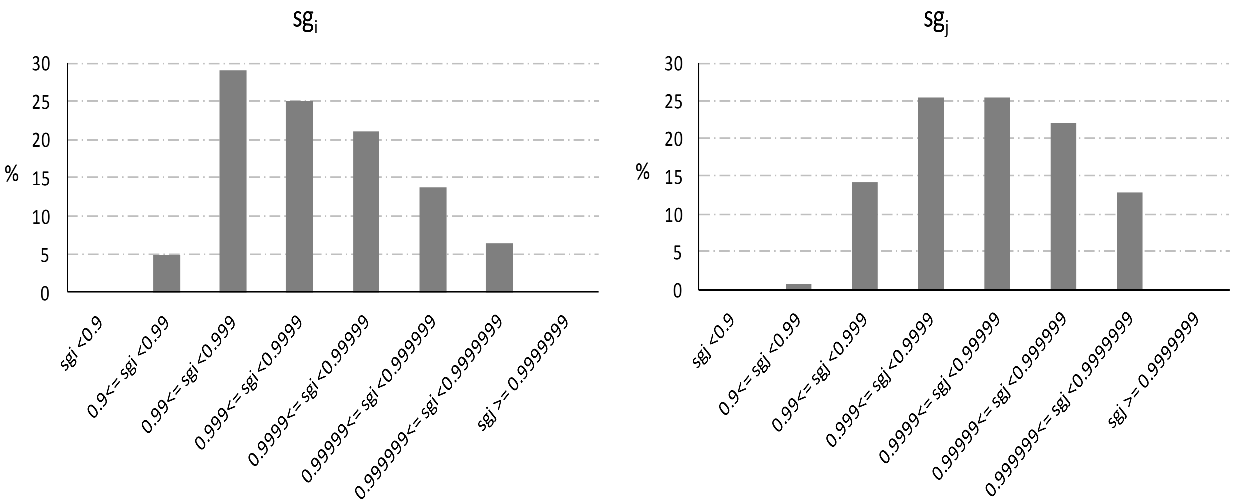


Fig. 3. Distributions of segmentations.

[16] With a sufficiently large sample, matches can be found for even the most unlikely combinations, in which case the propensity to match is very low but non-zero. This zero-frequency problem can be treated with some type of smoothing (e.g., Laplace or add-one smoothing). See, for example, Liu (2011).

3. Clusters

In the previous section, the workers, the jobs, and the worker-job matches have been divided into groups according to their characteristics, resulting in a large number of groups, which may be difficult to interpret. Now, using a clustering methodology, we can reduce the number of groups to a smaller number of clusters – *worker clusters, job clusters,* and *joint clusters* (formed by the joint groups corresponding to the matches of workers in a given worker cluster with jobs in a given job cluster) – until a single cluster for the entire labour market is attained. We shall also show that the clustering process give us a better overview of the structure of the local labour markets. With symmetric characteristics, if we consider the same groupings on both sides, then each worker cluster has a mirror job cluster, in which the job groups will be the mirrors of the worker groups in the worker cluster, and vice versa; and we will also have mirror joint clusters in which the job cluster will be the mirror of the worker cluster. A single period for the whole time interval of the data is considered, so that subscript t can be ignored[17].

The clustering methodology must be based on a previously defined similarity measure. In the context of labour matching, we consider that worker (job) groups are more similar, the more they resemble in the way they match with job (worker) groups. Following this approach, we define similarity $sw_{i1\text{-}i2}$ between each pair of worker groups W_{i1}-W_{i2} as the overlapping or percentage of coincidence of the distribution of their effective probabilities $pm_{ij}\ p_j$ of matching with each of the different job groups j

$$sw_{i1\text{-}i2} = \sum_{j=1}^{m} \min\,(pm_{i1j}\ p_j,\ pm_{i2j}\ p_j\ ) \ = \sum_{j=1}^{m} p_j\ \ \min\,(pm_{i1j},\ pm_{i2j}\ ) \tag{5}$$

Its value will be between one (if the distributions are identical) and zero (if the job groups which match the workers of W_{i1} fail to coincide with any of the job groups which match the workers of W_{i2}).

We can define the similarity $sj_{j1\text{-}j2}$ between each pair of job groups J_{j1}-J_{j2} in an analogous way.

Using this concept of similarity, we can graphically show how the joint clusters with high propensity to match may be considered as labour market clusters. In Fig. 4a, we represent the joint groups as elements of a matrix in which the rows and columns represent worker and job groups, respectively. The darker colour in each element indicates a higher propensity to match of the corresponding joint group. With symmetric characteristics, the joint groups corresponding to the main diagonal would be mirror joint groups, generally with high propensity to match. In Fig. 4b, the worker groups that most resemble in the way they match with job groups are put together in worker clusters. Within each worker cluster, the elements of each job group (which form 'little columns') will have similar propensity to match, which is high in the darker 'little columns'. In Fig. 4c, the job groups that most resemble in the way they match with worker groups are put together in job clusters and the matrix is partitioned in blocks corresponding to the joint clusters. The elements of each joint cluster have similar propensity to match. The dark joint clusters in Fig. 4c correspond to joint clusters in which workers and jobs have high propensity to match and, in this sense, may be considered as labour market clusters. In the most extreme case, in which the labour market clusters are 'pure islands', the propensity to match outside these islands is zero.

[17] Obviously, we can repeat the cluster analysis for successive time intervals, which would allow us to study its evolution in this dimension.

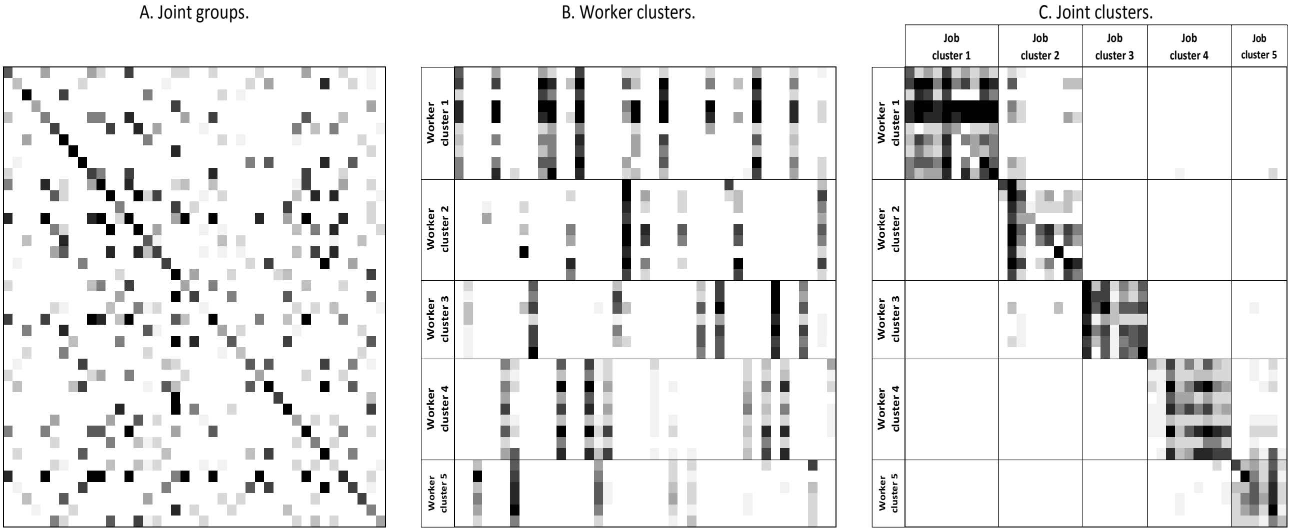


Fig. 4. Clustering process.

With symmetric characteristics, and if the pair of job groups J_{j1}-J_{j2} are the mirrors of the pair of worker groups W_{i1}-W_{i2}, the similarities $sw_{i1\text{-}i2}$ and $sj_{j1\text{-}j2}$ of the mirror pairs are highly positively correlated. The reason is that if $sw_{i1\text{-}i2}$ is high, then pm_{i1j} and pm_{i2j} will be high for the same job groups in most cases, and hence, by taking into account the high positive correlation of the propensities to match of symmetric joint groups, pm_{ij1} and pm_{ij2} will also be high for the respective symmetric joint groups (i=j). Therefore, $sj_{j1\text{-}j2}$ will be high too. Given the high positive correlation between the similarities of the mirror pairs, the clustering of the worker groups using the similarities between each pair of worker groups will be similar, but generally not identical, to the clustering of the job groups using the similarities between each pair of job groups. If we want to obtain the same grouping on both sides, the arithmetic mean can be used as a measure of similarity

$$s_{ij} = (sw_{ij} + sj_{ij}) / 2 \tag{6}$$

If we work with the same groupings on both sides, then the joint clusters are square blocks and those on the main diagonal are mirror joint clusters. The joint groups of each mirror joint cluster will have high propensity to match, similar to the elements in the main diagonal belonging to the mirror joint cluster. Therefore, in this case, the labour market clusters will be located principally on the main diagonal. In the extreme case of 'pure islands', we have a block diagonal matrix.

We use a hierarchical method of clustering, with groups gradually fusing to form increasingly larger groups. This method starts by merging the two groups with the highest similarity into a new group or cluster; the similarity of this new group with the rest of the groups is then recalculated, and the next two groups with the highest similarity are merged together. This process continues until we obtain a single cluster for the entire labour market. It can be visualised with a graphical display called a dendrogram or tree diagram. The process can be stopped when a specified number of clusters is reached or when the highest similarity falls below a specified level[18].

[18] Cotterman and Peracchi (1992) propose a methodology to identify optimal groupings of industries by minimising a loss function that combines goodness-of-fit and parsimony in the estimation of a wage equation.

By applying the methodology described to our data, we encounter computational problems related to the large size of our database, and additional problems due to the insufficient quantity of information for certain groups to be statistically representative. In order to overcome those problems, we have selected a sample of 1,587 common groups that appear in the 10,000 joint groups with the most matches.

In Fig. 5, we show, considering a single four-year period, the distribution of similarities for pairs of worker and job groups in this sample. As might be expected, almost all pairs show very low similarity and only a very small percentage show high similarity. A clear positive correlation (0.79) is also found between the similarities of the mirror pairs.

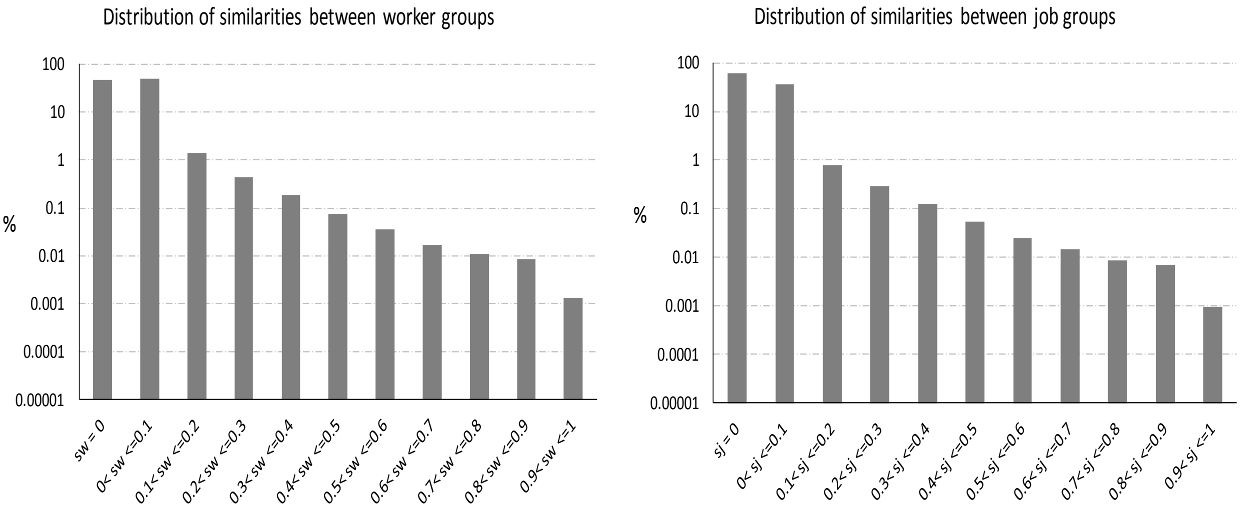


Fig. 5. Distribution of similarities.

By following the hierarchical method described above[19], we have developed the clustering of the 1,587 groups, with the same groupings on both sides, until a single cluster is obtained. An overview of the structure of the labour market that the clustering shows is reflected in Fig. 6, equivalent to Fig. 4c. In the figure, the sixteen clusters of the Andalusian labour market that are detailed in Table A1 in the Appendix are marked. We observe a high degree of segmentation[20], as illustrated by the dark joint clusters around the main diagonal, but we also observe that these clusters are not 'pure islands', as shown by the dark elements (reflecting high propensity to match) outside these clusters. It should be borne in mind that there are 'less frequent' matches not included in this clustering, which constitute an area to be explored.

[19] The calculations have been performed with STATA. To reduce computational effort to reasonable limits, we have used the correlation between the distributions of effective probabilities as a proxy for the measure of similarity. We have also used the cluster average-linkage utility instead of recalculating the similarity of the newly formed groups with the rest of the groups. With smaller databases, we have checked that our results are hardly affected by the use of these approximations.

[20] Álvarez de Toledo et al. (2008, 2011) test, with macroeconomic and individual data from the Spanish public employment agencies respectively, the plausibility of the stock-flow model (Coles and Smith, 1998) for the Spanish economy. In essence, they conclude that there exists clear evidence of this type of labour market segmentation. More specifically, the results point to an extreme case of that scheme: a queue of workers.

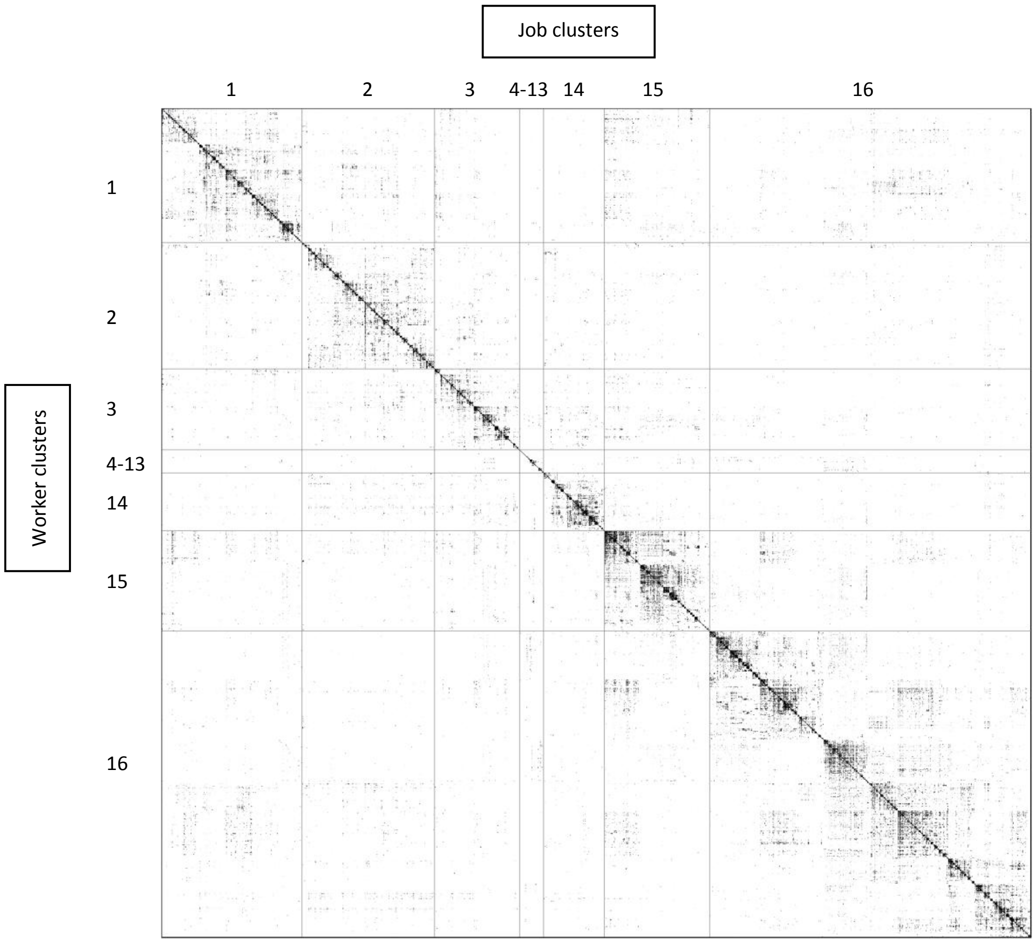


Fig. 6. Cluster structure of the labour market.

As explained above, the clustering process can be carried out until a specified number of clusters is obtained. In addition to the sixteen clusters represented in Fig. 6, in Section 5 we will also use a grouping in six 'big clusters'. As an example, we also show a dendrogram of 25 groups[21] in Fig. A1 in the Appendix.

4. Mobility and willingness to move

We consider symmetric characteristics and, again, a single period for the whole time interval of data, so that the subscript t[22] can be ignored.

If a worker of the worker group W_i matches a job of his mirror job group $J_{j=i}$, this means that the set of job characteristics corresponds exactly to the set of worker characteristics and, therefore, the mobility in this match can be interpreted as minimal. If this worker matches a job of another job group $J_{j\neq i}$, we will say that mobility is low if the worker group W_i has a high similarity with W_j, the mirror worker group of J_j – we can consider that the worker has to move from W_i to W_j before matching –.

[21] We do not show the complete dendrogram for the 1,587 groups due to lack of space.

[22] As we previously mentioned, we could repeat the exercise for successive time intervals, which would allow us to study the temporal evolution.

Following this approach, we define the measure of mobility mb_{ij} in the matches of the joint group S_{ij} as:

$$mb_{ij} = 1 - sw_{i\text{-}j} \tag{7}$$

whose value is between zero and one[23].

A multiple regression model can now be estimated, whose dependent variable is the measure of mobility in each worker-job match; that variable is presented as a function of worker and job characteristics and macroeconomic conditions. Table 1 presents the results of the estimation. We estimate three specifications which differ in the number of covariates and therefore in the number of observations (matches), since some covariates have missing observations in our sample. Specification (1) includes all the covariates under consideration, while the other two specifications fail to take certain covariates into account.

We obtain a reasonable positive effect on mobility of the tightness (or *V/U* ratio = θ)[24] in the job group and a negative effect of the tightness in the worker group[25]. A logical positive effect on mobility of the previous mobility shown by the worker is also observed, which has been obtained as the mean of the observed mobilities in the previous matches. We find a negative effect of the job group segmentation and of the worker group segmentation[26]. The negative effects from both variables could be explained because more segmentation in the worker or in the job group means further isolation, which should result in lower mobility.

The rest of the variables included in the specifications are more conventional and their effect differs little from previous literature. We can highlight the negative effect of search duration on mobility[27]. Unemployment benefits affect mobility adversely[28]. Men[29], young workers, non-nationals, those with less education and those belonging to qualified manual occupations[30], show a higher mobility in relative terms. Previous studies differ from ours in these last results, since they generally conclude that better educated or highly qualified workers show greater geographical or sectorial/occupational mobility. In our case, this could be explained by the forced mobility outside the construction sector of many workers with low education and qualification due to the sharp crisis in the Spanish housing market during our period of study.

[23] Notice that this variable defines worker mobility in a broad sense, considering jointly geographical, occupational and sectorial mobility. Other authors usually do not consider these three mobilities jointly, therefore our results are comparable to those of them only up to some extent. In this sense, we can mention some papers that combine at least two of the mentioned mobilities; for example, Elliott and Lindley (2006) analyse occupational and sectorial mobility in the Italian economy, and McQuaid (2006) analyses the occupational and spatial mobility in some Scottish regions.

[24] The vacancy-unemployment ratio for each group has been obtained by rescaling the SAE administrative stocks of vacancies and unemployed workers using information about the outflows: we use the rescaling factor given by the ratio "total job placements / matches involving registered job offers in the SAE" in the case of the vacancies, and the ratio "total job placements / matches involving registered workers in the SAE" in the case of the unemployed workers.

[25] Ahn et al. (1999) observe that the vacancy rate in the departure region decreases migration willingness among males.

[26] Except in specification (1).

[27] However, Ahn et al. (1999) do not observe any significant effect of unemployment duration on inter-regional migration willingness.

[28] Antolin and Bover (1997) also observe a lower propensity to mobility by those workers enrolled in the public employment agencies.

[29] Except in specification (1).

[30] Particularly, in specification (3).

Table 1. Regression model of mobility with standard errors adjusted for groups of workers.

Dependent variable: mb_{ij} (mobility)			(1)	(2)	(3)
		Constant	0.182*** *(0.005)*	0.186*** *(0.005)*	0.124*** *(0.002)*
Job characteristics	Segmentation of the job group	ln (sg_j)	-3.951*** *(0.075)*		
	Tightness of the job group	ln (θ_j)	0.044*** *(0.001)*		
Worker characteristics	Segmentation of the worker group	ln (sg_i)	1.545*** *(0.098)*	-1.215*** *(0.076)*	-1.561*** *(0.036)*
	Tightness of the worker group	ln (θ_i)	-0.026*** *(0.001)*	-0.008*** *(0)*	
		Previous mobility	0.561*** *(0.002)*	0.621*** *(0.002)*	0.761*** *(0.001)*
		Choose several occupations	0.015*** *(0.001)*	0.013*** *(0.001)*	0.006*** *(0.001)*
	Duration of the search (ref: 0 - 90 days)	91 - 180 days	*0.001* *(0.001)*	*-0.003** *(0.001)*	
		181 - 365 days	*-0.002* *(0.002)*	*-0.011**** *(0.002)*	
		366 or more days	*-0.010*** *(0.004)*	*-0.018**** *(0.004)*	
	Gender (ref: male)	Female	0.002 *(0.002)*	-0.006*** *(0.001)*	-0.005*** *(0.001)*
	Age (ref: 16-29 years old)	30-44 years old	-0.018*** *(0.002)*	-0.018*** *(0.002)*	-0.012*** *(0.001)*
		45-54 years old	-0.030*** *(0.002)*	-0.029*** *(0.002)*	-0.020*** *(0.001)*
		55 years old or more	-0.043*** *(0.003)*	-0.045*** *(0.002)*	-0.031*** *(0.001)*
	Nationality (ref: outside EU)	Spanish	-0.013*** *(0.003)*	-0.017*** *(0.002)*	-0.014*** *(0.001)*
		EU	-0.011 *(0.006)*	-0.005 *(0.006)*	-0.006 *(0.004)*
	Education (ref: secondary education (general))	Illiterate/No education	0.001 *(0.004)*	0.007 *(0.004)*	0.007*** *(0.001)*
		Primary education	-0.003* *(0.002)*	0 *(0.002)*	0.003*** *(0.001)*
		Secondary education (vocational training programmes)	0.005 *(0.003)*	0 *(0.003)*	-0.001 *(0.002)*
		Postsecondary (professional technicians)	0.01 *(0.006)*	0 *(0.005)*	-0.001 *(0.003)*
		Postsecondary (university and others)	0.002 *(0.008)*	-0.011 *(0.007)*	0.007 *(0.004)*
	Sector of activity (ref: contruction)	Agriculture	-0.126*** *(0.003)*	-0.081*** *(0.002)*	-0.043*** *(0.001)*
		Industry	-0.084*** *(0.005)*	-0.073*** *(0.005)*	-0.038*** *(0.004)*
		Trade, catering, transport, communications	-0.065*** *(0.003)*	-0.048*** *(0.003)*	-0.034*** *(0.002)*
		Financial services, business services	(omitted)	(omitted)	-0.021 *(0.036)*
		Public services	-0.089*** *(0.004)*	-0.054*** *(0.003)*	-0.031*** *(0.002)*
		Other services	-0.051*** *(0.003)*	-0.043*** *(0.003)*	-0.044*** *(0.002)*
	Group of occupation (ref: qualified manual worker)	Qualified non-manual worker	-0.008 *(0.01)*	-0.030** *(0.009)*	-0.039*** *(0.005)*
		Non-qualified non-manual worker	0.028*** *(0.003)*	0 *(0.003)*	-0.008*** *(0.002)*
		Non-qualified manual worker	0.017*** *(0.002)*	0.005*** *(0.002)*	-0.014*** *(0.001)*
	Province (ref: Seville)	Almeria	0.056*** *(0.003)*	-0.016*** *(0.003)*	0.007*** *(0.002)*
		Cadiz	0.048*** *(0.002)*	0.013*** *(0.002)*	0.019*** *(0.001)*
		Cordoba	0.008** *(0.002)*	0 *(0.002)*	-0.001 *(0.001)*
		Granada	0.053*** *(0.003)*	0.009*** *(0.003)*	0.001 *(0.001)*
		Huelva	0.003 *(0.003)*	0.005 *(0.003)*	0.012*** *(0.001)*
		Jaen	0.008** *(0.003)*	-0.004 *(0.003)*	0.001 *(0.001)*
		Malaga	0.038*** *(0.002)*	-0.002 *(0.002)*	0.005*** *(0.001)*
	Administrative status (ref: registered unemployed)	Employed	-0.009*** *(0.002)*	0.013*** *(0.002)*	
		TEAS (subsidised temporary agricultural workers)	-0.014*** *(0.003)*	0.013*** *(0.002)*	
		Other categories	0.001 *(0.003)*	0.009** *(0.003)*	
	Search scope (ref: searching beyond the region)	Searching only within the municipality	0.004 *(0.004)*	-0.004 *(0.003)*	
		Searching only within the province	-0.001 *(0.002)*	-0.007*** *(0.002)*	
		Searching only within the region	0.010* *(0.004)*	0.008* *(0.004)*	
		Unemployment benefit	-0.007*** *(0.001)*	-0.002 *(0.001)*	
Macroeconomic conditions		EPA unemployment rate	-0.001*** *(0)*	-0.000** *(0)*	
		Number of observations	292,107	347,901	1,493,792
		R2(adjusted)	38.4%	34.3%	45.8%

Coefficient (standard error).

* p<0.05; ** p<0.01; *** p<0.001

We use the results from specification (3) to obtain a measure of the 'a priori' workers' willingness to move before they match, which will be used in the next section. This specification only considers information about worker characteristics. It contains fewer variables than specification (2), but a larger number of observations is available.

5. Unemployment duration

We now show the usefulness of some of the tools that we have proposed in the previous sections, by including them in an unemployment duration model along with other conventional variables. In order to obtain an inflow sample, we have selected, from among the observations (matches) used as input data in Section 3, those corresponding to registered workers in the SAE whose date of registration occurs within the sample period, thereby obtaining a sample of 1,003,927 matches corresponding to 357,053 different workers (roughly, 3 spells per worker on average). Among other variables, we include the segmentation of the worker group and the worker's willingness to move, as measured in the previous sections. Our measure of mobility is also used to define different destination states.

A lognormal duration model with multiple exits, recurrent events and shared frailty is estimated, where the exits (or matches) can be divided into four types[31] depending on the mobility in the match: '*nm*' are matches with no mobility (belonging to a mirror joint group); '*lm*' are matches with low mobility ($0 < mb_{ij} \leq 0.4$); '*mm*' are matches with medium mobility ($0.4 < mb_{ij} \leq 0.95$); and '*hm*' are matches with high mobility ($mb_{ij} > 0.95$).

Table 2 presents the results of the estimation. To begin with, higher worker group segmentation generally increases unemployment duration with the main exception of matches with no mobility, which seems reasonable if one considers that a highly isolated group experiences no congestion from the arrival of seekers coming from other groups.

The interpretation of the effect of the willingness to move is more complicated, although it also seems reasonable. Generally, a greater willingness to move (from low to medium levels of willingness) reduces unemployment duration by only a certain extent. In matches with high mobility, higher levels of willingness to move further reduce unemployment duration, while in matches with no mobility, a greater willingness to move increases unemployment duration. The results obtained for the variable 'Choose several occupations' (workers who declare themselves willing to work in various possible occupations when they are registered at the employment office, unlike those who only manifest one desired occupation) are similar to those obtained for the willingness to move.

A high level of tightness in the worker group reduces unemployment duration in matches with no mobility, but in matches with mobility the effect is rather the opposite: intuition tells us that a low *V/U* ratio supposes a stimulus to move to other groups, especially to those closest. Overall, the negative effect prevails.

[31] On the technique of competing risk duration models with recurrent events and shared frailty see Cameron and Trivedi (2009) and Cleves et al. (2010). Durations of more than 2 years are treated as censored at 2 years, due to their relatively small number of observations.

Table 2. Estimation of a lognormal duration model with multiple failure and shared frailty for the hazard rate of the jobseekers.

	Covariates	Total exits	No mobility (nm)	Low mobility (lm)	Medium mobility (mm)	High mobility (hm)
Segmentation of the worker group (sg_i) (ref: low)	Medium	0.165***	-0.058***	0.428***	-0.074***	0.175***
		(0.005)	*(0.008)*	*(0.009)*	*(0.013)*	*(0.011)*
	High	0.169***	-0.272***	0.558***	0.088***	0.362***
		(0.007)	*(0.011)*	*(0.013)*	*(0.018)*	*(0.015)*
Willingness to move (ref: low)	Medium	-0.266***	0.816***	-1.279***	-1.573***	-1.421***
		(0.006)	*(0.009)*	*(0.012)*	*(0.017)*	*(0.015)*
	High	-0.079***	1.787***	-0.062***	-1.579***	-2.109***
		(0.008)	*(0.016)*	*(0.018)*	*(0.022)*	*(0.018)*
	Choose several occupations	-0.006	0.228***	-0.134***	-0.115***	-0.146***
		(0.004)	*(0.007)*	*(0.008)*	*(0.011)*	*(0.009)*
Tightness of the worker group (θ_i) (ref: low)	Medium	-0.037***	-0.369***	0.238***	0.208***	0.016
		(0.007)	*(0.012)*	*(0.014)*	*(0.019)*	*(0.016)*
	High	-0.127***	-0.480***	0.532***	-0.078**	-0.038
		(0.009)	*(0.016)*	*(0.02)*	*(0.026)*	*(0.022)*
Six big clusters (ref: 1)	2	0.376***	0.565***	0.115***	0.678***	0.686***
		(0.015)	*(0.027)*	*(0.028)*	*(0.045)*	*(0.034)*
	3	-0.654***	-0.600***	-0.769***	0.786***	0.281**
		(0.034)	*(0.052)*	*(0.081)*	*(0.12)*	*(0.091)*
	4	0.670***	0.605***	0.810***	0.727***	0.995***
		(0.035)	*(0.058)*	*(0.068)*	*(0.102)*	*(0.09)*
	5	-0.243***	0.275***	-0.832***	-0.417***	-0.05
		(0.015)	*(0.025)*	*(0.031)*	*(0.045)*	*(0.035)*
	6	-0.091***	0.356***	-0.893***	-0.501***	0.232***
		(0.011)	*(0.018)*	*(0.023)*	*(0.03)*	*(0.023)*
Gender (ref: male)	Female	0.349***	0.345***	0.286***	0.293***	0.657***
		(0.005)	*(0.008)*	*(0.009)*	*(0.013)*	*(0.011)*
Age (ref: 16-29 years old)	30-44 years old	-0.014**	-0.070***	-0.047***	0.097***	0.122***
		(0.005)	*(0.009)*	*(0.01)*	*(0.013)*	*(0.011)*
	45-54 years old	0.023***	-0.128***	-0.002	0.176***	0.312***
		(0.006)	*(0.01)*	*(0.012)*	*(0.016)*	*(0.014)*
	55 years old or more	0.231***	0.000	0.220***	0.416***	0.721***
		(0.008)	*(0.014)*	*(0.016)*	*(0.023)*	*(0.019)*
Nationality (ref: outside EU)	Spanish	-0.228***	-0.253***	-0.166***	-0.166***	-0.164***
		(0.007)	*(0.013)*	*(0.015)*	*(0.02)*	*(0.017)*
	EU	-0.086***	-0.120***	-0.021	-0.125**	0.067
		(0.018)	*(0.031)*	*(0.036)*	*(0.046)*	*(0.044)*
Education (ref: secondary education (general))	Illiterate/No education	-0.246***	-0.107***	-0.408***	-0.098***	-0.210***
		(0.01)	*(0.018)*	*(0.02)*	*(0.028)*	*(0.024)*
	Primary education	0.019***	0.064***	-0.006	-0.018	-0.019
		(0.005)	*(0.009)*	*(0.01)*	*(0.013)*	*(0.011)*
	Secondary education (vocational training programmes)	-0.055***	-0.125***	0.083***	0.018	-0.101***
		(0.011)	*(0.02)*	*(0.022)*	*(0.029)*	*(0.026)*
	Postsecondary (professional technicians)	0.023	-0.086**	0.106**	0.058	-0.088
		(0.019)	*(0.033)*	*(0.038)*	*(0.049)*	*(0.045)*
	Postsecondary (university and others)	-0.144***	-0.176***	-0.004	-0.118	-0.419***
		(0.025)	*(0.044)*	*(0.052)*	*(0.066)*	*(0.056)*
Sector of activity (ref: contruction)	Agriculture	-0.552***	-0.516***	-0.787***	-0.460***	0.094***
		(0.007)	*(0.012)*	*(0.014)*	*(0.019)*	*(0.016)*
	Industry	-0.116***	-0.207***	-0.051	-0.353***	0.861***
		(0.025)	*(0.042)*	*(0.052)*	*(0.072)*	*(0.063)*
	Trade, catering, transport, communications	-0.457***	-0.556***	-0.450***	-0.914***	0.422***
		(0.009)	*(0.017)*	*(0.018)*	*(0.025)*	*(0.022)*
	Financial services, business services	-0.304	-0.703*	0	-0.205	0.000
		(0.186)	*(0.321)*	*()*	*(0.552)*	*(0.359)*
	Public services	-0.599***	-0.532***	-0.872***	-0.340***	0.025
		(0.013)	*(0.022)*	*(0.024)*	*(0.034)*	*(0.029)*
	Other services	-0.672***	-0.367***	-1.125***	-1.032***	0.343***
		(0.01)	*(0.018)*	*(0.02)*	*(0.027)*	*(0.023)*
Group of occupation (ref: qualified manual worker)	Qualified non-manual worker	0.222***	-0.077	0.106	0.246*	0.490***
		(0.036)	*(0.06)*	*(0.078)*	*(0.109)*	*(0.088)*
	Non-qualified non-manual worker	0.128***	-0.080***	0.555***	-0.225***	0.151***
		(0.009)	*(0.017)*	*(0.018)*	*(0.024)*	*(0.022)*
	Non-qualified manual worker	0.001	-0.358***	0.247***	-0.061***	-0.071***
		(0.005)	*(0.01)*	*(0.01)*	*(0.015)*	*(0.012)*
Province (ref: Seville)	Almeria	-0.011	0.395***	-1.034***	-0.450***	0.273**
		(0.037)	*(0.062)*	*(0.073)*	*(0.108)*	*(0.095)*
	Cadiz	0.167***	0.259***	0.185***	0.201***	0.105***
		(0.007)	*(0.013)*	*(0.015)*	*(0.02)*	*(0.016)*
	Cordoba	-0.049***	0.312***	-1.133***	-0.130***	0.268***
		(0.011)	*(0.019)*	*(0.023)*	*(0.032)*	*(0.025)*
	Granada	-0.213***	0.026	-1.207***	-0.439***	0.348***
		(0.019)	*(0.033)*	*(0.037)*	*(0.055)*	*(0.042)*
	Huelva	0.290***	0.486***	-0.097***	0.278***	0.736***
		(0.009)	*(0.016)*	*(0.017)*	*(0.023)*	*(0.02)*
	Jaen	0.099***	0.602***	-1.190***	0.051	0.433***
		(0.012)	*(0.021)*	*(0.025)*	*(0.034)*	*(0.027)*
	Malaga	0.526***	0.509***	0.079*	0.464***	0.727***
		(0.017)	*(0.028)*	*(0.035)*	*(0.049)*	*(0.038)*
	Unemployment benefit	0.379***	0.408***	0.382***	0.375***	0.379***
		(0.004)	*(0.006)*	*(0.007)*	*(0.009)*	*(0.008)*
Macroeconomic conditions	EPA unemployment rate	0.004***	0.004***	0.004***	0.018***	0.010***
		(0)	*(0.001)*	*(0.001)*	*(0.001)*	*(0.001)*
	Sigma	1.181***	1.550***	1.591***	1.874***	1.761***
		(0.001)	*(0.001)*	*(0.002)*	*(0.002)*	*(0.002)*
	Theta	0.250***	1.049***	1.510***	2.731***	1.718***
		(0.005)	*(0.007)*	*(0.006)*	*(0.007)*	*(0.008)*
	Log-likelihood	-982,790	-623,456	-542,528	-384,566	-446,585
	Number of observations	1,003,927	1,003,927	1,003,927	1,003,927	1,003,927
	Number of subjects	357,053	357,053	357,053	357,053	357,053
	Number of transitions	1,001,263	344,425	303,613	156,056	197,169
	LR chi2 test (Prob > chi2)	0.00	0.00	0.00	0.00	0.00
	LR test of theta = 0 (Prob>=chibar2)	0.00	0.00	0.00	0.00	0.00

Coefficient (standard error).
legend: * p<0.05; ** p<0.01; *** p<0.001

To jointly control for the three characteristics which define our groups (municipality, group of occupation, and sector of economic activity), we have considered a grouping into six 'big clusters', following the methodology explained in Section 3. The shortest unemployment durations correspond to the big cluster 3, which is mainly located in the provincial capitals of Seville, Malaga and Granada and the village of Ubrique (in Cadiz), whose predominant occupations include those of nurses, skilled workers in agriculture, and leatherwork artisans or similar. On the other hand, the longest unemployment durations correspond to the big cluster 4, which is largely located in Almeria, with the provincial capital of Almeria, El Ejido and Roquetas de Mar as the most representative municipalities, and with agriculture, manufacturing industries, and construction as predominant sectors.

No significantly different effects are obtained for other variables commonly included in previous estimates. Women generally present longer unemployment duration than men, particularly in matches of high mobility. In general, the shortest unemployment duration corresponds to workers between 30 and 44 years old and the longest duration to workers of 55 years old or more, however, the results are not uniform for matches with different mobility. For instance, the youngest workers (16-29 years) have the shortest unemployment duration in matches with medium and high mobility. National workers present shorter unemployment duration than do foreigners[32], particularly in matches of no mobility. With regard to education, the workers in the two extremes (illiterate/no education and university[33] and other postsecondary non-technicians) have shorter unemployment duration than those with intermediate levels. Low reservation wages for the unskilled workers and a higher rate of job offers for the most skilled workers could be behind these results. Workers in construction experience the longest unemployment duration, which is plausible considering that the temporal period of our data coincides with the Spanish housing crisis. However, in matches of high mobility, the workers from construction present the shortest unemployment duration and the workers from industry the longest duration. The results according to occupation level (qualified/non-qualified, manual/non-manual) vary with the types of mobility without showing a clear pattern. In terms of provinces, Malaga, Huelva and Cadiz, in this order, experience longer unemployment duration, but the effect for other provinces vary considerably with the types of mobility. Unemployment benefit recipients clearly experience longer unemployment duration. Finally, the unemployment rate has, in general, a small but positive and significant effect on unemployment duration, which implies a slight procyclical hazard rate[34].

Figure 7 represents the estimated individual hazards. Once we control for (observed and unobserved) heterogeneity, we notice, for all hazard rates, an increase in the first days of search (probably due to administrative factors), and a later gradual decline; this decline indicates that workers tend to become more dependent on being unemployed over time[35]. Several reasons may explain this pattern: among others, the search intensity may decrease, the general and specific

[32] Carrasco and García-Pérez (2008) state that immigrants could show a higher probability of leaving unemployment than natives if we do not control for unobserved heterogeneity.

[33] Bover et al. (2002) find that a university degree increases the hazard of leaving unemployment only during the first 3 months; afterwards the hazard reduces to levels below those of less educated workers. These findings are consistent with the high incidence of long-term unemployment among highly educated unemployed individuals (Machin and Manning, 1999).

[34] However, Antolin (1997) finds a possible countercyclical or acyclical behaviour for the hazard rate.

[35] This result has also been observed by other authors (van den Berg and van Ours (1996), Shimer (2007, 2008) and, for the Spanish economy, Uña-Álvarez et al. (2003), Güell and Hu (2006)) but it remains inconclusive in the literature. Other authors, such as Machin and Manning (1999), Steiner (2001), and Ebrahimy and Shimer (2010), find no clear decline in the hazard rate.

skills of the unemployed workers are progressively lost, or there is a stigma effect that makes those remaining in unemployment less attractive for employers. Hazard ratios reveal that the hazard decline with duration is slightly more pronounced in matches with no or low mobility than in matches with medium or high mobility.

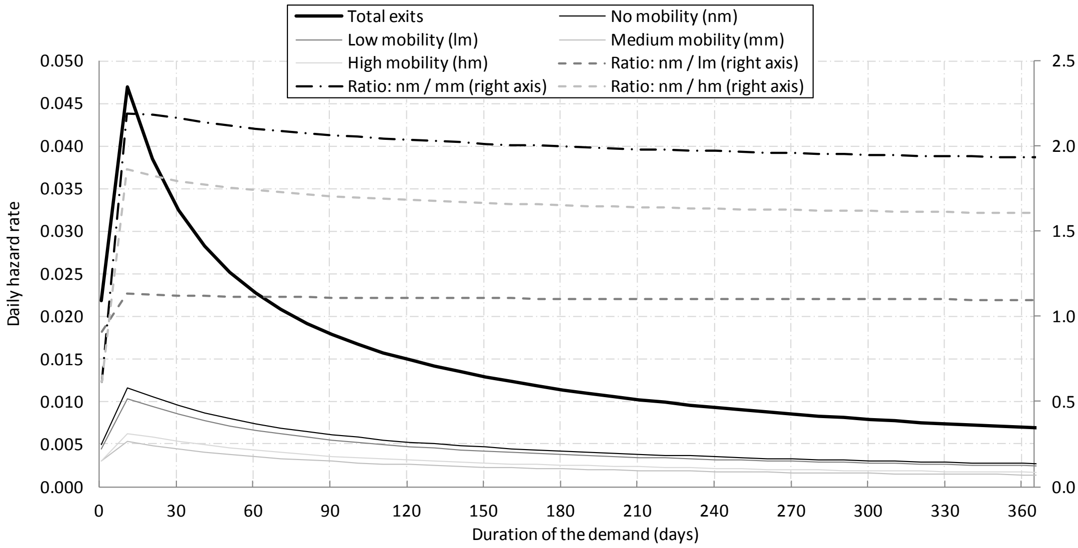


Fig. 7. Estimated hazard rates (and ratios) for a job demand depending on the exit.

In principle, a worker seeking work globally (for all types of mobility) could accumulate the corresponding hazard rates, but this idea requires further investigation in terms of whether the search efficiency in each type could diminish.

6. Conclusions

In this paper we propose empirical tools to account for the role of heterogeneities in search and matching theories applied to labour economics, and we show an application to the Andalusian labour market, by using a large database of individual microdata.

We have analysed the concept of labour market segmentation and proposed empirical measures related to this concept: propensity to match, and segmentation in worker and job groups. The results of our application show a high degree of segmentation. We use a clustering methodology, based on a similarity measure, to attain a better overview of the structure of the labour market and to reduce the large number of worker and job groups to a manageable number of clusters. We show in which way the worker-job clusters with high propensity to match that are formed may be considered labour market clusters. The clustering again highlights a high degree of segmentation, which is reflected in labour market clusters with high propensity to match, but these clusters are not 'pure islands', as shown by the existence of worker-job groups with high propensity to match outside these clusters.

A measure of mobility in each worker-job match is proposed, directly related to our similarity measure, and a regression model is then estimated that relates mobility to worker and job characteristics and to macroeconomic conditions. Mobility is higher, the higher the worker mobility in previous matches, the lower the segmentation of the worker or job group, the lower the tightness in the worker group, or the higher the tightness in the job group. With few exceptions, no significantly different effect is obtained from other variables commonly included in studies in this field. The results of the regression model are used to estimate the 'a priori' workers' willingness to move.

We show the usefulness of the tools that we have developed by including them in an unemployment duration model along with other conventional variables. The unemployment duration is higher for workers with lower willingness to move, and for those with higher segmentation or with lower tightness within their worker group. These overall results may change when we take into account the different types of exits. For example, lower willingness to move and higher segmentation in the worker group, which implies less competition from external workers, reduces unemployment duration for matches in the worker's own group (matches without mobility). Unemployment duration is also lower for workers with lower tightness in their worker group when matches take place outside the worker's own group (matches with mobility), which indicates that these workers experience a clear incentive to move. The hazard rate of the worker tends to fall with duration except in the first days of search. Again, no significantly different effects are obtained for other variables commonly included in previous literature in this field.

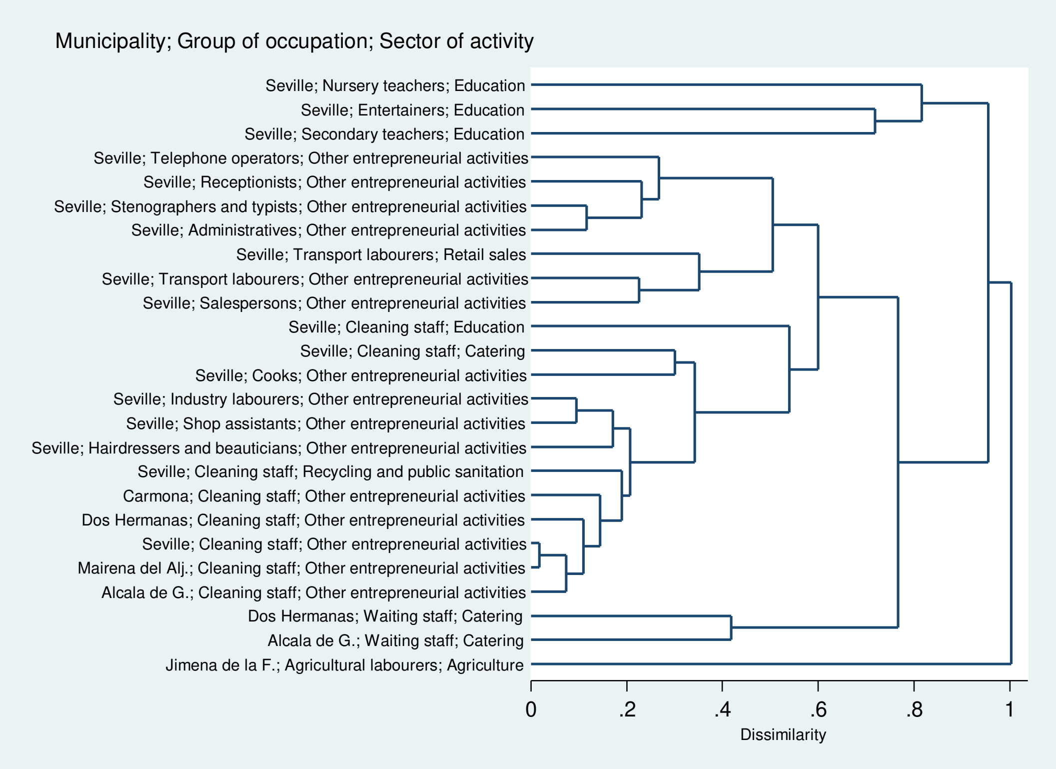
Worker mobility, geographical or occupational, and the availability of relevant information are important requirements for effective labour matching, and constitute a prominent element that should be taken into account to guide the design of active labour market policies. The empirical tools proposed in this paper may be useful in this regard, by helping jobseekers and firms looking for workers to follow successful paths previously used by others. The clustering methodology allows past information on matches to be processed in order to generate a 'roadmap' of possible routes to different labour market clusters, which can also include the probability of success in each route. The versatility of the methodology proposed makes it possible to enrich the information provided from this perspective and to take into consideration other variables of interest, such as the best search channels for each cluster. Further research is required to test the practical usefulness of this methodology for real labour intermediation.

Appendix

Table A1. The Andalusian labour market structured into 16 clusters.

Cluster	Matches	Province	Province (%)	Municipality	Province of the municipality	Municipality (%)	Group of occupation	Occupation (%)	Sector of activity	Sector (%)
1	419189	Cordoba	74.0	Cordoba	Cordoba	10.2	Other qualified workers in agriculture	52.4	Agriculture	81.9
		Seville	17.7	Palma del Rio	Cordoba	8.6	Agricultural labourers	30.3	Construction	9.5
		Malaga	6.0	Aguilar de la Frontera	Cordoba	6.1	Bricklayers	5.3	Other entrepreneurial activities	2.9
		Jaen	1.8	Baena	Cordoba	4.2	Construction labourers	2.3	Others	
		Cadiz	0.5	Nueva Carteya	Cordoba	3.6	Shop assistants	2.2		
				Bujalance	Cordoba	3.5	Cleaning staff	2.1		
				Cabra	Cordoba	3.1	Others			
				Osuna	Seville	2.9				
				Ecija	Seville	2.9				
				Castro del Rio	Cordoba	2.3				
				Montalban de Cordoba	Cordoba	2.3				
				Posadas	Cordoba	2.2				
				Cañete de las Torres	Cordoba	2.0				
				Doña Mencia	Cordoba	2.0				
				Montilla	Cordoba	2.0				
				Others						
2	269366	Jaen	94.6	Jodar	Jaen	9.2	Other qualified workers in agriculture	74.3	Agriculture	84.0
		Granada	3.0	Porcuna	Jaen	4.4	Agricultural labourers	12.6	Construction	9.0
		Cadiz	1.3	Jaen	Jaen	4.4	Bricklayers	5.2	Public Administration	2.4
		Cordoba	0.6	Villacarrillo	Jaen	4.3	Construction labourers	2.1	Others	
		Seville	0.3	Alcaudete	Jaen	3.5	Others			
		Malaga	0.2	Navas de San Juan	Jaen	3.5				
				Andujar	Jaen	3.4				
				Martos	Jaen	3.4				
				Santisteban del Puerto	Jaen	3.3				
				Jamilena	Jaen	2.7				
				Villanueva del Arzobisp	Jaen	2.6				
				Arjonilla	Jaen	2.3				
				Marmolejo	Jaen	2.2				
				Arjona	Jaen	2.0				
				Castellar	Jaen	2.0				
				Others						
3	143737	Granada	81.4	Granada	Granada	17.3	Other qualified workers in agriculture	64.8	Agriculture	62.0
		Jaen	9.7	Motril	Granada	11.5	Agricultural labourers	7.5	Construction	13.2
		Cordoba	6.8	Alcala la Real	Jaen	8.0	Cleaning staff	4.7	Other entrepreneurial activities	6.4
		Malaga	2.2	Illora	Granada	7.0	Bricklayers	4.5	Catering	6.2
				Loja	Granada	6.6	Shop assistants	4.0	Public Administration	6.1
				Huetor Tajar	Granada	6.6	Waiting staff	3.8	Retail sales	2.8
				Montefrio	Granada	5.6	Construction labourers	3.2	Others	
				Priego de Cordoba	Cordoba	5.4	Others			
				Pinos Puente	Granada	4.0				
				Almuñecar	Granada	3.0				
				Salar	Granada	2.3				
				Iznalloz	Granada	2.1				
				Moclin	Granada	2.0				
				Others						
4	376	Granada	100.0	Motril	Granada	100.0	Operators of cranes and similar machinery	100.0	Maritime transport	77.1
									Travel agencies	22.9
5	1028	Malaga	50.5	Lanjaron	Granada	22.8	Other qualified workers in agriculture	50.5	Construction	50.5
		Granada	30.2	Cartajima	Malaga	11.8	Industry labourers	22.8	Food and drinks	22.8
		Cadiz	9.9	Farajan	Malaga	10.9	Machinery Mechanics	9.9	Health activities and social services	16.8
		Jaen	9.4	Salares	Malaga	10.6	Nursing assistants	9.4	Sale and repair of motor vehicles	9.9
				Archez	Malaga	10.1	Cleaning staff	7.4		
				Algeciras	Cadiz	9.9				
				Porcuna	Jaen	9.4				
				Guadix	Granada	7.4				
				Alpandeire	Malaga	7.1				
6	168	Huelva	100.0	Huelva	Huelva	100.0	Postal staff	100.0	Post and telecommunications	100.0
7	286	Badajoz	100.0	Olivenza	Badajoz	100.0	Other qualified workers in agriculture	100.0	Agriculture	100.0
8	1215	Cadiz	100.0	Ubrique	Cadiz	60.2	Artisans	60.2	Leather and footwear	60.2
				Jerez de la Frontera	Cadiz	24.0	Post	33.5	Post and telecommunications	33.5
				Chiclana de la Frontera	Cadiz	9.5	Construction labourers	6.3	Construction	6.3
				Villaluenga del Rosario	Cadiz	6.3				
9	330	Granada	73.3	Granada	Granada	73.3	Pharmacists	100.0	Retail sales	100.0
		Malaga	26.7	Malaga	Malaga	26.7				
10	4230	Malaga	33.9	Malaga	Malaga	26.3	Nurses	98.0	Health activities and social services	100.0
		Cadiz	25.7	Almeria	Almeria	10.3	Nursing assistants	2.0		
		Almeria	15.5	Jerez de la Frontera	Cadiz	8.7				
		Seville	8.2	Seville	Seville	8.2				
		Granada	7.7	Granada	Granada	7.7				
		Huelva	4.9	Marbella	Malaga	7.5				
		Jaen	2.1	Cadiz	Cadiz	7.2				
		Others	2.0	El Ejido	Almeria	5.2				
				Huelva	Huelva	4.9				
				El Puerto de Santa Mari	Cadiz	4.9				
				Sanlucar de Barrameda	Cadiz	4.9				
				Linares	Jaen	2.1				
				Others		2.0				
11	239	Seville	70.3	Seville	Seville	36.0	Nursing assistants	64.0	Health activities and social services	100.0
		Jaen	29.7	Ecija	Seville	34.3	Physiotherapists	36.0		
				Vilches	Jaen	29.7				
12	1427	Seville	47.2	Seville	Seville	47.2	Photographers	53.6	Leisure activities	93.8
		Madrid	24.1	Madrid	Madrid	24.1	Actors and directors	40.2	Health activities and social services	6.2
		Granada	23.2	Granada	Granada	23.2	Doctors	6.2		
		Badajoz	5.5	Merida	Badajoz	5.5				
13	354	Huelva	100.0	Valverde del Camino	Huelva	100.0	Shoemakers	56.5	Leather and footwear	100.0
							Operators in footwear manufacture	43.5		
14	83267	Almeria	93.5	Almeria	Almeria	26.6	Agricultural labourers	37.6	Agriculture	44.0
		Granada	5.9	Ejido (El)	Almeria	22.8	Industry labourers	11.5	Construction	18.0
		Jaen	0.3	Roquetas de Mar	Almeria	18.1	Construction labourers	8.1	Wholesale sales	11.3
		Others	0.3	Nijar	Almeria	7.6	Bricklayers	7.4	Catering	9.8
				Vicar	Almeria	3.8	Waiting staff	5.9	Other entrepreneurial activities	6.2
				Adra	Almeria	2.3	Shop assistants	5.9	Retail sales	4.5
				Others			Other qualified workers in agriculture	5.5	Terrestrial transport	3.5
							Cleaning staff	5.1	Others	
							Truck Drivers	3.5		
							Transport Labourers	2.2		
							Others			
15	202129	Malaga	99.3	Malaga	Malaga	53.5	Cleaning staff	22.3	Catering	29.3
		Cadiz	0.4	Marbella	Malaga	10.3	Shop assistants	17.4	Construction	27.8
		Seville	0.3	Torremolinos	Malaga	4.5	Waiting staff	13.3	Other entrepreneurial activities	22.0
				Fuengirola	Malaga	4.5	Bricklayers	11.7	Retail sales	11.9
				Estepona	Malaga	4.1	Construction labourers	8.0	Health activities and social services	2.0
				Mijas	Malaga	4.0	Cooks	6.2	Others	
				Benalmadena	Malaga	3.8	Industry labourers	3.3		
				Velez-Malaga	Malaga	3.7	Other qualified workers in agriculture	2.6		
				Nerja	Malaga	2.3	Transport Labourers	2.5		
				Others			Others			
16	691141	Seville	49.6	Seville	Seville	13.6	Other qualified workers in agriculture	27.6	Agriculture	47.8
		Cadiz	31.2	Jerez de la Frontera	Cadiz	8.1	Agricultural labourers	21.9	Construction	21.8
		Huelva	19.1	Sanlucar de Barrameda	Cadiz	4.7	Bricklayers	9.4	Other entrepreneurial activities	11.1
		Others	0.1	Palacios y Villafranca (L	Seville	3.7	Cleaning staff	8.9	Catering	8.2
				Almonte	Huelva	3.4	Shop assistants	7.9	Retail sales	5.3
				Carmona	Seville	3.1	Construction labourers	6.2	Public administration	2.2
				Dos Hermanas	Seville	3.0	Waiting staff	4.2	Others	
				Tocina	Seville	2.7	Others			
				Cadiz	Cadiz	2.5				
				Arcos de la Frontera	Cadiz	2.3				
				Huelva	Huelva	2.2				
				Lebrija	Seville	2.1				
				Utrera	Seville	2.0				
				Others						

Fig. A1. Example of dendrogram for 25 groups.



References

Abraham, K.G., Katz, L.F., 1986. Cyclical unemployment. Sectoral shifts or aggregate disturbances? Journal of Political Economy. 94 (3), 507-522.

Ahn, N., de la Rica, S., Ugidos, A., 1999. Willingness to move for work and unemployment duration in Spain. Economica. 66 (263), 335-357.

Alba-Ramírez, A., 1994. Self-employment in the midst of unemployment. The case of Spain and the United States. Applied Economics. 26 (3), 189-204.

Álvarez de Toledo, P., Núñez, F., Usabiaga, C., 2008. La función de emparejamiento y el mercado de trabajo español. Revista de Economía Aplicada. 16 (48), 5-35.

Álvarez de Toledo, P., Núñez, F., Usabiaga, C., 2011. An empirical analysis of the matching process in Andalusian public employment agencies. Hacienda Pública Española. 198 (3), 67-102.

Álvarez de Toledo, P., Núñez, F., Usabiaga, C., 2012. Segmentación laboral y análisis de clusters con datos individuales. Una aplicación al mercado de trabajo andaluz. Centro de Estudios Andaluces. Documento de Trabajo E2012/03.

Antolín, P., 1997. Los flujos de trabajadores en el mercado de trabajo español. Papeles de Economía Española. 72, 154-167.

Antolín, P., Bover, O., 1997. Regional migration in Spain: The effects of personal characteristics and of unemployment, wage and house price differentials using pooled cross-sections. Oxford Bulletin of Economics and Statistics. 59 (2), 215-235.

Arellano, M., Bover, O., 2002. Learning about migration decisions from the migrants: Using complementary datasets to model intra-regional migrations in Spain. Journal of Population Economics. 152 (2), 357-380.

Bachmann, R., Burda, M.C., 2010. Sectoral transformation, turbulence and labor market dynamics in Germany. German Economic Review. 11 (1), 37-59.

Barnichon, R., Figura, A., 2011. Labor market heterogeneities, matching efficiency, and the cyclical behavior of the job finding rate. CREI. Mimeo.

Bentolila, S., Dolado, J.J., Jimeno, J.F., 2012. Reforming an insider-outsider labor market: The Spanish experience. FEDEA. Documento de Trabajo 2012-01.

Bover, O., Arellano, M., Bentolila, S., 2002. Unemployment duration, benefit duration, and the business cycle. Economic Journal. 112 (479), 223-265.

Brown, A.J.G., Christian, M., Snower, D.J., 2009. An incentive theory of matching. CEPR. Discussion Paper 7283.

Cameron, A.C., Trivedi, P.K., 2009. Microeconometrics: Methods and applications. Cambridge University Press, Cambridge.

Carrasco, R. 1999. Transitions to and from self-employment in Spain: An empirical analysis. Oxford Bulletin of Economics and Statistics. 61 (3), 315-341.

Carrasco, R., García-Pérez, J.I., 2008. Unemployment duration among immigrants and natives. Unobserved heterogeneity in a multi-spell duration model. Universidad Carlos III, Economic Series (33). Working Paper 08-69.

Cotterman, R., Peracchi, F., 1992. Classification and aggregation: An application to industrial classification in CPS data. Journal of Applied Econometrics. 7 (1), 31-51.

Cleves, M., Gould, W., Gutierrez, R.G., Marchenko, Y.V., 2010. An introduction to survival analysis using stata, third ed. Stata Press.

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

Devine, T.J., Kiefer, N.M., 1991. Empirical labor economics: The search approach. Oxford University Press, Oxford.

Ebrahimy, E., Shimer, R., 2010. Stock-flow matching. Journal of Economic Theory. 145 (4), 1325-1353.
Elliott, R.J.R., Lindley, J., 2006. Skill specificity and labour mobility: Occupational and sectoral dimensions. Manchester School. 74 (3), 389-413.
Faini, R., Galli, G., Gennari, P., Rossi, F. 1997. An empirical puzzle: Falling migration and growing unemployment differentials among Italian regions. European Economic Review. 41 (3-5), 571-579.
Fox, J.T., 2008. Estimating matching games with transfers. NBER. Working Paper 14382.
Gautier, P.A., 2002. Non-sequential search, screening externalities and the public good role of recruitment offices. Economic Modelling. 19 (2), 179-196
Greenwood, M.J., 1985. Human migration: Theory, models, and empirical studies. Journal of Regional Science. 25 (4), 521-544.
Güell, M., Hu, L., 2006. Estimating the probability of leaving unemployment using uncompleted spells from repeated cross-section data. Journal of Econometrics. 133 (1), 307-341.
Güell, M., Petrongolo, B., 2007. How binding are legal limits? Transitions from temporary to permanent work in Spain. Labour Economics. 14 (2), 153-183.
Iglesias-Fernández, C., Llorente-Heras, R., 2007. Sectoral structure, qualification characteristics and patterns of labour mobility. The Service Industries Journal. 27 (4), 411-434.
Jain, A.K., Murty, M.N., Flynn, P.J., 1999. Data clustering: A review. ACM Computing Surveys. 31 (3), 265-323.
Jovanovic, B., Moffitt, R., 1990. An estimate of a sectoral model of labor mobility. Journal of Political Economy. 98 (4), 827-852.
Kambourov, G., Manovskii, I., 2009. Occupational mobility and wage inequality. Review of Economic Studies. 76 (2), 731-759.
Kuo, M., Smith, E., 2009. Marketplace matching in Britain: Evidence from individual unemployment spells. Labour Economics. 16 (1), 37-46.
Lagos, R., 2000. An alternative approach to search frictions. Journal of Political Economy. 108 (5), 851-873.
Lalé, E., 2012. Trends in occupational mobility in France: 1982-2009. Labour Economics. 19 (3), 373-387.
Lancaster, T., 1979. Econometric models for the duration of unemployment. Econometrica. 47 (4), 939-956.
Lilien, D.M., 1982. Sectoral shifts and cyclical unemployment. Journal of Political Economy. 90 (4), 777-793.
Liu, B., 2011. Web data mining: Exploring hyperlinks, contents, and usage data, second ed. Springer-Verlag, Berlin.
Lucas, R., Prescott, E.C., 1974. Equilibrium search and unemployment. Journal of Economic Theory. 7(2), 188-209.
Machin, S., Manning, A., 1999. The causes and consequences of longterm unemployment in Europe, in: Ashenfelter, O., Card, D. (Eds), Handbook of Labor Economics 3C. North-Holland, Amsterdam, pp. 3085-3139.
McCall, B.P., 1990. Occupational matching: A test of sorts. Journal of Political Economy. 98 (1), 45-69.
McQuaid, R., 2006. Job search success and employability in local labor markets. The Annals of Regional Science. 40 (2), 407-421.
Miller, R.A., 1984. Job matching and occupational choice. The Journal of Political Economy. 92 (6), 1086-1120.
Moreno-Galbis, E., 2009. Comparing the effect of labor market institutions on employment dynamics. Annales d'Economie et de Statistique. 95-96, 141-166.

Mortensen, D.T., 2009. Island matching. Journal of Economic Theory. 144 (6), 2336-2353.
Mortensen, D.T., Pissarides, C.A., 1999. New developments in models of search in the labor market, in: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics 3B. North-Holland, Amsterdam, pp. 2567-2627.
Neal, D., 1995. Industry-specific human capital: Evidence from displaced workers. Journal of Labor Economics. 13 (4), 653-677.
Petrongolo, B., Pissarides, C.A., 2001. Looking into the black box: A survey of the matching function. Journal of Economic Literature. 39 (2), 390-431.
Phelps, E.S. (Ed.), 1970. Microeconomic foundations of employment and inflation theory. W.W. Norton, New York.
Pissarides, C.A., 2000. Equilibrium unemployment theory. The MIT Press, Cambridge (Mass.).
Pissarides, C.A., 2008. EconomicDynamics Interviews Christopher Pissarides on the Matching Function. EconomicDynamics Newsletter & Mailing List. 10 (1).
Pissarides, C.A., Wadsworth, J., 1989. Unemployment and the inter-regional mobility of labour. The Economic Journal. 99 (397), 739-755.
Reich, M., Gordon, D.M., Edwards, R.C., 1973. A theory of labor market segmentation. American Economic Review. 63 (2), 359-365.
Rogerson, R., Shimer, R., 2011. Search in macroeconomic models of the labor market, in: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics 4A. North-Holland, Amsterdam, pp. 619-700.
Rogerson, R., Shimer, R., Wright, R., 2005. Search-theoretic models of the labor market: A survey. Journal of Economic Literature. 43 (4), 959-988.
Sattinger, M., 2010. Queueing and searching. University at Albany-SUNY, Department of Economics. Mimeo.
Shimer, R., 2007. Mismatch. American Economic Review. 97 (4), 1074-1101.
Shimer, R., 2008. The probability of finding a Job. American Economic Review. 98 (2), 268-273.
Steiner, V., 2001. Unemployment persistence in the West German labour market. Negative duration dependence or sorting? Oxford Bulletin of Economics and Statistics. 63 (1), 91-113.
Topel, R.H., Ward, M.P., 1992. Job mobility and the careers of young men. Quarterly Journal of Economics. 107 (2), 439-479.
Uña-ÁLvarez, J., Otero-Giráldez, M.S., Álvarez-Llorente, G., 2003. Estimation under length-bias and right-censoring: An application to unemployment duration analysis for married women. Journal of Applied Statistics. 30 (3), 283-291.
van den Berg, G.J., van Ours, J.C., 1996. Unemployment dynamics and duration dependence. Journal of Labor Economics. 14 (1), 100-125.
Yashiv, E., 2007. Labor search and matching in macroeconomics. European Economic Review. 51 (8), 1859-1895.