WORKING PAPER

Labour market change

Labour market segmentation: Piloting new empirical analyses – Methodology

[Labour market segmentation: Piloting a new quantitative and policy analysis]

Authors: Irene Cruz, Oscar Molina and Joan Miquel Verd (QUIT); Carlos Vacas- Soriano (Eurofound)
Research Managers: Carlos Vacas-Soriano and Irene Mandl

Eurofound reference number: WPEF19012

Related report/s: Labour market segmentation: Piloting new empirical and policy analyses

The European Foundation for the Improvement of Living and Working Conditions (Eurofound) is a tripartite European Union Agency established in 1975. Its role is to provide knowledge in the area of social, employment and work-related policies according to Regulation (EU) 2019/127.

European Foundation for the Improvement of Living and Working Conditions

Telephone: (+353 1) 204 31 00

Email: information@eurofound.europa.eu

Web: www.eurofound.europa.eu

Contents

This working paper presents the detailed methodology used in the empirical analysis of the report *Labour market segmentation: Piloting new empirical and policy analyses* (which is mainly presented in Chapter 3). The purpose of this methodological Annex is to provide a detailed account of the way in which the empirical analysis has been operationalised and implemented in the four countries studied. More specifically, it starts by introducing the different steps of the quantitative analysis and then it explains which variables have been included in the analysis of careers and how they have been used in each country. Moreover, it also shows how the alphabets have been constructed and how they have been implemented in the four countries.

Box 1. Glossary of methodological terms

Labour market status – relation between an individual and the labour market, namely whether they are employed, unemployed.

Labour market states – more specifically, the type of legal implementation of labour market status, either through specific contractual arrangements (for example full-time open ended) or cause of leave (maternity, family care or leave of absence in aggregate form).

Labour market career trajectory groups – groups of individuals per country, based on experiencing similar labour market trajectories, as resulting from the analysis of career sequences in the four selected countries.

Cross-sectional analysis –analysis of observations for many respondents in the same moment in time.

Longitudinal analysis – analysis of series of observations obtained from many respondents over time, also referred to as panel analysis.

Alphabet - series or streams of pre-defined states that in this study serve to build sequences.

Alphabet A – the baseline alphabet used in this study, based on the type of contract and working time.

Alphabet B – the alphabet used in this study whenever possible (not in France), based on type of labour market contract, working time, occupational category and earnings.

Sequences – series of ordered pre-defined states.

Optimal matching – statistical method aimed at finding patterns of similarity / dissimilarity between sequences.

Standardisation index – index that measures the distance between an individual labour market career and the standard labour market career characterised by full time stable employment and high earnings.

Entropy – an index that measures how difficult it is to predict the next state in an individual's career.

Study approach and steps followed

The empirical study conducted in the report proposes an innovative approach to cover the analysis of labour market segmentation in four selected countries: France, Germany, Spain and the UK. These countries have been selected because they have available and ready to use longitudinal data and they represent different institutional configurations and labour market problems.

This quantitative study consisted of working with national experts in France, Germany, Spain and UK, who were in charge of handling the data and running statistical testing on it. Meanwhile, the methodological approach was being developed throughout the quantitative analysis by the core research team, who based on incremental outputs produced by national experts were regularly selecting the tools/statistical tests most relevant for answering the specific research questions under study here. The questions include the following: what are the labour market structures in each of these counties? Which labour market career trajectory groups emerge? Who is affected and which are the drivers and effects of labour market segmentation?

The quantitative analysis consisted of six steps and included different types of analyses, as illustrated in Figure 1 below. To briefly summarise, it draws on longitudinal data in four countries to first of all consider the full range of information available on individuals in relation to the labour market in each country and provides a static picture (steps 1 and 2). The next step (step 3) in the analysis marks already the start of the use of a longitudinal approach and serves to explore labour market transitions, and what characteristics are conducive to mobility versus entrapment in specific employment situations. This is followed by building a comparative scenario that matches different types of careers based on their similarity or proximity, and as such allows to explore differences between these trajectory groups (step 4). Based on the analysis of similarity / dissimilarity, step 5 consists of constructing combined indicators that reflect a person's career, and includes the distance to the standard (open-ended, full -time job), and a variable on entropy which indicates the 'uncertainty' or possibility of predicting one's career trajectory. A final phase (step 6) then consists of the analysis of what factors from the labour supply-side (age, gender) and labour demand-side (like sector and company size) are most important in predicting the likelihood for individuals to approach the standard career scenario. Each of the steps are presented in detail below.

Figure 1: Step-by-step quantitative analysis

Step 1
Identifying, accsessing and preparing the data;

Step 2
Building the sequences: defining states and Alphabets
2.1 Selecting states; 2.2. Construct Alphabet A; 2.3 Construct Alphabet B;

Step 3
Calculating transition rates
3.1.Transitions between states in Alphabet A; 3.2 Transitions between states in Alphabet B;

Step 4
Optimal matching, standartisation and clusters
4.1 Apply optimal matching OMA to sequences; 4.2. Build standardisation index; 4.3. Apply cluster analysis on sequences;

Step 5
Building synthetic career indicators
5.1 Standardisation; 5.2 Entropy; 5.3 Turbulence; 5.4 Complexity; 5.5 Volatility;

Step 6
Regression analysis
6.1.On transition probabilities; 6.2 On career standardisation index;

Source: Authors

Step 1: Identifying, accessing and preparing data

The first and vital step for the possibility of conducting the type of quantitative analysis presented here is access to data that provide information on the same individuals and their relation to the labour market over time - longitudinal data.

Identifying. Identification of suitable datasets for longitudinal analysis required considerable consultation with experts on the ground in specific countries in relation to aspects like characteristics of the data, waiting times for access it and similar. Importantly, the identified data sources include a mix of administrative and survey data (see Box 2) with administrative data used in Spain, and survey data in Germany, France and the UK. Key in relation to the different types of data used are the limitations to comparing results from the analysis in each country. Key differences across datasets are presented in Box 2 below.

The use of different types of datasets has important implications for the interpretation of results, particularly from a comparative perspective. More specifically, even though a similar analytical strategy has been implemented, the results will never be fully comparable due to the differences in the data used. The aim of the analysis has been to make results as comparable as possible, not to achieve full comparison, as this is impossible with the datasets included.

Important for the interpretation of results in relation to the types of data used is the implicit bias in panel structure-type surveys, especially when they are on households as is the case in Germany and the UK, towards an oversampling of individuals that may be better off. Administrative data account for everybody rather than people who respond to surveys, likely including more disadvantaged population than possible to reach in a panel (which may exclude for example individuals in prisons).

Accessing. For the identified data sources, the study team applied for access in the cases (France and the UK) where the data was not already available for use by the involved research teams. In cases where there was a formal application process, access was granted in a period of up to two months.

Preparing. Data preparation was handled by national experts with guidance from the core research team. National experts cleaned the data to leave only the variables used in this analysis and prepared a matrix comprising as detailed information as was available in each country on the different events in relation to the labour market experienced by individuals in the sample. See Box 2 below on the main differences in relation to the data sources used in the quantitative analysis.

→ **Step 1 results**. The study team identified four longitudinal datasets suitable for the quantitative analysis, prepared and cleaned the data keeping only needed variables and organised it in a way to facilitate the building of sequences as described in the second step below.

Box 2. Types of data used in the quantitative analysis

France. Survey data: Formation et Qualification Professionnelle (FQP) 2009-2014.

- Information on pay available at beginning and end points in the data, not throughout the period under observation (this also determined that analysis of careers in France is limited to changes in relation to labour market states (see Table 4 based only on contracts rather than also pay and occupational category as in the other three countries – see Table 5).

- No information on company size;

- No information on working time.

Germany. Survey data: German Socio-Economic Panel (GSOEP) 2001-2008; 2009-2016.

- In order to avoid the bias caused by sample attrition over a 15-year period, the sample has been split into two periods (see Table 6). As such comparison is not meaningful and differences between two periods are not exploited in the analysis;

- Almost no observations for individuals under the age of 25 after preparing the data for the analysis;

- Limited observations for individuals under the age of 35 after preparing the data for the analysis;

- Inactive people not included in the sample;

- No information on leave of absence.

Spain. Administrative data: Continuous Sample of Working Lives (MCVL) 2000-2008; 2009-2016.

- As the only administrative data source of the countries under study, includes 'uninformed' periods in individual careers where there are gaps in information on specific individuals;

- Uninformed periods are treated as a residual variable that corresponds to two possible states: unemployed not receiving benefits or inactive, with no disaggregated information on these two states available;

- Disaggregated information on internship/training contracts available (not the case in the other three countries).

UK. Survey data British Household Panel Survey (BHPS) 2001-2008.

- Disaggregated information available and used on more detailed labour market states, including: student; sick/disabled and different forms of leave – maternity, family care;

- No meaningful information on foreign-born individuals in the sample;

- No wage information, meaning a proxy on financial security was used in its place.

Step 2: Building the sequences: defining states and alphabets

The use of different datasets that cover different variables and periods, rely on different sample sizes and moreover draw on household panels in some countries and administrative sources in others, was a key challenge in this study. This has required flexibility in order to adapt the same logic of the quantitative analysis to the different characteristics of the longitudinal datasets used whilst maintaining the maximum degree of comparability. The way in which this adaptation was made is described in more detail in the next section on the alphabets.

Defining states. The first step in building the sequences consists in identifying a list of possible states in relation to the labour market, including based on the three proxies used to group workers into trajectory groups (contractual arrangements, earnings and occupation). Datasets varied in terms of the richness they were able to provide, and as such presented different opportunities in terms of depth of analysis (see Box 2). For this reason, the research team adopted the use of two 'alphabets', namely a finite set of pre-defined possible states.

Defining alphabets. Next in this step was **building alphabet A**, which can be considered the baseline alphabet. It covers basic contract-related variables which are available in all datasets (see Table 1). The two main types of information it provides is type of contract and cause of leave. The type of contract can range from the 'standard' contract that is full-time open-ended, to part-time, fixed-term, self-employment or internship. Cause of leave-related states include leave of absence, retirement, unemployment, or being completely outside of the labour force without benefits or inactive. Importantly, the full range of states differs across country depending on the richness of data (see alphabet A in the next section).

The final stage in step 2 was to **construct alphabet B,** which is more comprehensive compared to alphabet A, as it includes also information on earnings (counted by hourly pay), weekly working hours, and professional category (see alphabet B in the next section). This more elaborate alphabet was possible to apply in all countries except France because earnings information was available only at the beginning and end of the observation period rather than for the whole sequence as in other countries. Alphabet B is thus a multidimensional typology of states, that are ranked from best to worst-off in alphabetical order. State A therefore is the one that identifies open-ended, full-time work, matched with high pay and high skill professional categories. The subsequent states represent situations departing proportionally from this standard employment scenario. These resulting states do not represent a specific contractual arrangement, but rather rich multidimensional information simplified into ranked states.

Drawing from the two alphabets (A and B), this step is continued by building the sequences or careers. Sequences are defined as ordered lists of pre-defined moments on a time axis. In the UK, Spain and Germany, sequences are built based on alphabet B, whereas in France this is done based on alphabet A. The sequences are then used for the analysis starting in Step 4.

→ **Results Step 2.** Step two resulted in the operationalisation of labour market states and two alphabets. The analysis of labour market structures constitutes the building blocks in the analysis of transitions (described in Step 3 below). This part of the analysis is cross-sectional, meaning that the distribution of individuals in relation to the different states is shown in specific points in time, rather than throughout their careers (as this part of the analysis begins with Step 4).

Step 3: Calculating transition rates

This step consists in considering transition rates between states for both of the alphabets, and as such represents the first consideration of the longitudinal perspective. Here, the analysis looks at how often individuals in the sample transition between any two given states. As such, it considers both the direction of the transitions, for example from temporary to permanent jobs or vice versa, and the frequency of transitions. Rates for such transitions are then calculated by comparing the actual transitions against permanence in a given state. In this way, it is possible to observe which transitions are more frequent.

→ **Results Step 3**. Step 3 resulted in the calculation of transition rates between states considered in the two alphabets, a measure of the probability associated to each type of transition (ascending, horizontal, or descending), which can be interpreted in terms of un/permeability of the states.

Step 4: Optimal matching, standardisation and clusters

This step represents a shift of focus to the whole career span of individuals in the different samples. Based on the sequences built in Step 2, optimal matching analysis (OMA) is applied to obtain a synthetic indicator of similarity/dissimilarity of individual trajectories in comparison to the ideal-type standard career, as previously defined in this study. Using OMA allows to find the distance between every sequence to this standard, and as such, to summarise what is potentially a very complex sequence pattern with just one synthetic variable, namely the degree of career standardization. The values of this variables after rescaling it range from 0 (complete standardization) to 1 (complete difference).

Because the ideal-type being used as a common reference across the four countries is very strict, it tends to overestimate non-standard careers, because every difference in relation to the standard trajectory (added or subtracted) is penalized. This fact needs to be taken into account when interpreting the results.

A final step here consists of applying cluster analysis to the standardization indexes, to ascertain whether the sequences fall into distinct types. In other words, by comparing all sequences against each other, it is possible to group those careers that are more similar among themselves. In this way, contrary to most LMS studies, this analysis allows to identify groups of individuals not only based on employment conditions they experience, but the sequences in which their careers unfold over time. These clusters, referred to in the results sections that follow as trajectory groups, should not strictly be interpreted as labour market segments although they somewhat approach that idea.

→ **Results Step 4.** Step 4 resulted in the identification of four clusters or career trajectory groups in each country based on the proximity or dissimilarity in their sequences. Moreover, it also delivered a synthetic standardisation index that summarises the distance between an individual's career and the standard career defined by stable, full time employment.

Step 5: Building synthetic career indicators

Synthetic career indicators allow collapsing very diverse information into a single quantifiable indicator by summarising the complexity of a given career into one variable, then allowing to order careers accordingly. One example of this synthetic career indicator is the aforementioned standardisation index, that summarises the distance between any career and the standard career. Another one is career entropy, or the degree of uncertainty or difficulty to predict sequences in an individual's career based on previously experienced trajectory.

→ **Results Step 5.** Step 5 resulted in the calculation of several synthetic career indicators, that provide information about uncertainty and volatility of careers. More specifically, an entropy index was computed that summarises the degree of uncertainty individuals face in their careers.

Step 6: Regression analysis

The last step in the analysis entails running Binomial Logistic Regression (see Box 3) on the standardisation index for all individual-level supply and demand-side factors under consideration, including age, occupation, main sector in the trajectory. Regression analysis aims to provide information about the role of these characteristics of individuals or their employment in explaining the individuals falling in one or another trajectory group. The distribution of the dependent variable (the standardisation index) tends to be polarized, so it was dichotomized for the analysis.

Box 3. Binomial Logistic Regression

A binomial logistic regression is used to predict the occurrence of an outcome (dependent variable) given the presence of some factors (independent variables). The dependent variable takes two values, 0 and 1. The model is used to predict the likelihood of different outcomes of the dependent variable. The outcome of the regression is a table with coefficients for each category of the independent variables, that should be interpreted in relation to the reference category used. More specifically, it allows interpreting a coefficient as the likelihood of outcome *X* changes, compared to the reference outcome, with a change in an independent variable.

→ **Results Step 6**. Step 6 resulted in a binomial logistic regression in each country providing information about the relationship between the standardisation index and a series of labour supply and demand-side variables. The coefficients in the regressions provide information about how statistically significant the variables included are.

Definition of states and alphabets

The study uses two alphabets in the analysis of careers, which are described in detail next. This section deals with the alphabets in a general way as based on their ideal implementation, while the next section explains in detail the specific way in which they required adaptations across the four selected countries.

Alphabet A

Alphabet A can be considered the baseline alphabet and covers basic contract-related variables which are available in all four countries analysed (see Table 1). More specifically, alphabet A considers the type of contract (open-ended versus temporary) and working time (full-time versus part-time). To make datasets comparable between countries, the first step was to distinguish four different types of labour market statuses for constructing careers based on alphabet A. The statuses are: employed, leave of absence, unemployed, inactive and other. At the same time, these labour market statuses contain within them different values or states.

Table 1: States considered for alphabet A

Labour Market Status	States considered	Variables	Notes on differences across countries
Employed	Full-time open ended	Type of contract	In France, open-ended includes both part-time and full-time, not disaggregated
	Full-time fixed-term		In France, fixed-term includes both part-time and full-time, not disaggregated
	Part-time open ended		
	Part-time fixed-term		
	Self-employed		
	Internship/ Training contracts		Only available for Spain
Leave of Absence	Leave of absence (general)*	Cause of leave	
	Maternity leave		Only available for UK
	Family care		Only available for UK
Unemployed	Unemployment with benefits		
	Unemployment without benefits		Not available in Spain (joint category with inactive – see uninformed)
	Inactivity	Inactivity	Not available in Germany; in Spain available as joint category with

Labour Market Status	States considered	Variables	Notes on differences across countries
Inactive / Out of the labour market			unemployed without benefits – see uninformed below.
	Full-time student		Only available for the UK
	Sick/disabled		Only available for the UK
	Uninformed / na	Uninformed period between two informed periods (including leaving the labour market temporarily)	In Spain, where administrative data includes periods unaccounted for, it is assumed for the analysis that uninformed is included as unemployed without benefits / inactive (unemployed and not seeking employment).

*Note: * leave of absence refers to a temporarily inactive period of time that an employee is away from the primary job. This can include paid or unpaid leave, whether for maternity/paternity or other care-related or other voluntary or involuntary leave.*

Source: Authors

Within the employed status, alphabet A includes six states. Four of them result from combining the type of contract (open-ended vs temporary) with working time (full-time vs part-time). The other two correspond to self-employment (when available in the dataset) and training contracts (when available in the dataset).

We also included leave of absence, referred to those situations where an employee is away from the primary job. This can include paid or unpaid leave, whether for maternity/paternity or other care-related or other voluntary or involuntary leave. Leave of absence refers to those situations that imply a temporary period of inactivity while maintaining the job.

The unemployed status comprises workers without a job but actively seeking one, independently of whether they receive any form of unemployment protection, whenever the dataset allows to distinguish the two states.

Meanwhile, the inactive category refers are those individuals that leave the labour market or are not active job seekers, for any reason, excluding those who retire or die over the period considered. This means that inactive people considered in the analysis are those that voluntarily (studies or taking care of family responsibilities) or involuntarily (unmotivated long-term unemployed) are out of the labour market. Finally, the 'uninformed' category includes individuals in the datasets – whether administrative (Spain) or survey data (UK, Germany, France) who either drop out of the sample (survey data) or have an incomplete employment record (administrative data). Only in the case of Spain, where the volume of uninformed periods is significantly higher compared to survey data, they've been considered for the analysis.

Country-specific adaptations of alphabet A

These states were adapted taking into consideration the availability of certain variables in national datasets and their specificities. The specific way in which the states in alphabet A have been adapted to the national datasets is shown in Table 2 and explained in more detail in the country sections in the following section. Table 2 contains two types of information. First, it shows which variables, corresponding to the states considered in alphabet A, are available in the datasets of the four countries. Secondly, in case the variable is not available, or is different, it details the information contained in the dataset.

Table 2: Availability of alphabet A states and adaptation to national datasets

Labour Market Status	States considered in alphabet A	Country			
		FR	DE	ES	UK
Employed	Full-time open ended	Open-ended	Yes	Yes	Yes
	Part-time open ended		Yes	Yes	Yes
	Full-time fixed-term	Fixed-term Temporary / Seasonal	Yes	Yes	Yes
	Part-time fixed-term		Yes	Yes	Yes
	Self-employed	Yes	Yes	Yes	Yes
	Internship/Training contracts	No	No	Yes	No
Leave of Absence	Leave of absence	Yes	No	Yes	Yes (Maternity leave)
Unemployed	Unemployment receiving benefits	Yes	Yes	Yes	Yes
	Unemployment not receiving benefits	Yes	Yes	Unemployed without benefits + Inactive	Yes
Inactive / Out of the labour market	Inactivity	Yes	No		Yes -Family care/home -Student -Sick/ Disabled
Other	Uninformed / N/a (Uninformed period between two informed periods (including leaving the labour market temporarily)	No	No		No

Source: Authors

As can be observed, the main differences between the datasets refer to the unemployed and inactive states. The different character of datasets (administrative vs survey data) together with the empirical difficulties to measure inactivity and distinguish it from unemployment, explain why more heterogeneity is found in these variables. In the case of Spain, whenever there is no information between two informed periods, or whenever information gaps are found in the trajectories, they can be due to inactivity or being unemployed but not receiving benefits: for this reason, a have a residual category is created consisting of all those who are either inactive or unemployed but do not perceive unemployment benefit.

Alphabet B

Compared to alphabet A, alphabet B is more comprehensive when it comes to the states, as it integrates other variables not considered in alphabet A. Moreover, states in alphabet B are built differently compared to alphabet A. Apart from the need for flexibility in the analysis, there are three other reasons for using alphabet B. First, it is important to include earnings (not part of alphabet A) as one of the key variables defining the state of the individual in the labour market. This is important for gauging the 'quality' behind specific contract types in different countries. Second, alphabet B provides a more accurate picture of the complexity associated with labour market structures, hence moving beyond an identification between type of contract and the segment. alphabet B thus captures the multidimensional character of LMS. Finally, the occupational category is a key variable explaining patterns of inequalities in the labour market and an individual's position in it.

Alphabet B is based on four variables:

Labour market status. The type of relation to the labour market is made up of the following categories: employed with a permanent contract; employed with a fixed-term or temporary contract; self-employed; unemployed; leave of absence; internship/training contract.

Working time. Part-time or working time coefficient is a variable that expresses the working hours as a proportion of a full-time job (40 hours per week). This variable was recoded into three categories: full-time (100%-85% of full-time), part-time (between 84% and 35% of full-time), and marginal-part-time (<35%).

Hourly pay. Earnings data and information vary significantly across countries. The objective in this case is to distinguish between low-pay, medium- or average-pay and high-paying jobs. In this vein, three pay categories were considered.

1. Earnings below 66% of the overall mean of medians in the period considered.
2. Earnings between 66% and 120% of the overall mean of medians in the period considered.
3. Earnings above 120% of the overall mean of medians in the period considered.

In each country considered, earnings data was adapted using the above categories.

Occupational category. Alphabet B also considers occupational category as a variable in order to capture the hierarchy in work relations as it reflects the skills level thus being representative of the job characteristics of individuals. There are three categories for this variable: high-skilled and higher hierarchy positions; medium-skilled or middle hierarchy positions; low-skilled categories. From the LMS point of view, professional categories are particularly important, because they are representative of the status a worker has in the labour market whilst also being an important driver explaining mobility patterns.

Alphabet B is used in all selected countries except France (Germany, Spain and UK). In order to obtain comparable results, the states have been built in all three countries according to the above-described definitions.

The variables in alphabet B are recoded into different categories (see Figure 2). These categories are later ranked and given a score. Every person, at any given point in time, has a total score resulting from the addition of the combination of these four ranked variables. The final score is finally transformed from integer to ordinal (into the letters composing alphabet B). Figure 2 represents the ideal approach to build alphabet B, although country deviations were necessary to adapt to the data limitations in each country (see next section).

There are two main reasons for building alphabet B in this way. First, by adding new variables and sub-categories, the number of combinations increases significantly. This allows to integrate different states that are similar from the point of view of an individual’s labour market status. Second, this allows to compare results among the three analysed countries (except France) at any point in time.

Columns in Figure 2 contain earnings and working time. For each of the states considered for these variables, a score is assigned. Rows contain the type of contract and occupational category. Again, for each of the states, a score is assigned. Each cell contains the sum of the scores corresponding to the four corresponding states.

Combining all the variables together yields a matrix containing 108 possible states. In order to collapse the complex information coming from these combinations, an aggregation mechanism was constructed for ranking all possible states. This was done by attributing a value for each of the combinations in the previous categorical variables. The rationale behind the ranking is the following:

- All variables are ordinal in the sense that they can be ranked according to the notion of better or worse working conditions. Better working conditions are given a higher score. So, if a variable has three categories, the best condition is assigned a score of three and the worst is assigned a score of one.
- Two variables were attached a greater weight: type of contract and hourly pay. The reason for assigning a greater weight to these two variables is because according to the existing literature, there is an agreement that these are the basic variables determining the labour market position of an individual. For these two variables, the best category has a score of four rather than three.

Figure 2: States considered for alphabet B

			Pay	< 66 %			66:120%			> 120%		
				1 1 1	1 1 1	1 1 1	2 2 2	2 2 2	2 2 2	4 4 4	4 4 4	4 4 4
		Working Time		full time	part-time	marginal pt	full time	part-time	marginal pt	full time	part-time	marginal pt
Labour Market Status		Occupational Category		3 3 3	2 2 2	1 1 1	3 3 3	2 2 2	1 1 1	3 3 3	2 2 2	1 1 1
Open-ended	4	upper tier	3	11	10	9	12	11	10	14	13	12
	4	mid tier	2	10	9	8	11	10	9	13	12	11
	4	lower tier	1	9	8	7	10	9	8	12	11	10
Self-Employed	2	upper tier	3	9	8	7	10	9	8	12	11	10
	2	mid tier	2	8	7	6	9	8	7	11	10	9
	2	lower tier	1	7	6	5	8	7	6	10	9	8
Temporary	2	upper tier	3	9	8	7	10	9	8	12	11	10
	2	mid tier	2	8	7	6	9	8	7	11	10	9
	2	lower tier	1	7	6	5	8	7	6	10	9	8
Internship / Apprenticeship	1	upper tier	3	8	7	6	9	8	7	11	10	9
	1	mid tier	2	7	6	5	8	7	6	10	9	8
	1	lower tier	1	6	5	4	7	6	5	9	8	7
Unemployment (receiving unemployment benefit)												
Leave of absence												
Unemployment (not receiving benefit) + Inactivity												

Note: Colours refer to the states defined in the alphabet. Numbers are the sum of scores for each combination of employment variables, where a higher score means better working conditions

Source: Authors

A	13 to 14
B	11 to 12
C	9 to 10
D	6 to 8
E	4 to 5
F	
G	

A total score was calculated for each combination of states by adding the score for each of the variables and categories. The result is a new numeric variable with a minimum value of four and a maximum value of 14 (unemployment and inactivity do not get a value). This is then recoded into an ordinal variable according to the states defined previously. Within each of the states in alphabet B there are several possible employment situations corresponding to different combinations of the four employment dimensions considered, which means it is impossible to assign a specific employment state to any of the states in alphabet B. However, Table 3 shows the employment profiles more likely to be found within each of the states.

Table 3: Predominant employment profiles for alphabet B

	Type of relationship	Part-time coefficient	Pay	Occupational status	Notes on differences across countries
State A	Permanent	Full -time	High-medium pay	Highly skilled	Corresponds to UK's State A and B
State B	Permanent, temporary	Full-time and part-time	High-medium pay	Highly skilled	Corresponds to UK's State C and D
State C	Self-employed, permanent, training/apprenticeship	Full-time and part-time	High-medium pay	Medium-low skilled	Corresponds to UK's State E and F
State D	Temporary, self-employed, training/apprenticeship	Part-time, marginal part-time	Medium-low pay	Medium-low skilled	Corresponds to UK's G (and partly H)
State E	Temporary, training/apprenticeship	Part-time, marginal part-time	Low pay	Low skilled	Corresponds to UK's H
State F	Unemployment (receiving unemployment benefit) and leave of absence				In UK includes unemployed, maternity leave, Student, Sick /disabled and Family care
State G	Unemployment (without unemployment benefits) and inactivity				

Source: Authors

It is important to note that these states should not be taken as predefined labour market segments. They simply represent an aggregation of states according to a rank ordering, which allows to explore the labour market structure and build the sequences or careers of individuals.

The differences in the information contained in the national datasets have required to introduce some country-specific adaptations. While a more detailed explanation of these country-specificities is provided in the next section, an overview of these cross-country adaptations is presented in Table 4:

- The most important differences are related to the non-employed states, F and G. State F represents those workers that are unemployed but benefit from income protection schemes, either contributory or assistance-based. Moreover, they include workers that are on leave (therefore not working) but that may return to their jobs once their leave period is finished. State G refers to workers that are unemployed and do not receive any form of unemployment protection or are simply inactive (not actively seeking a job). As will be explained in the country sections, the methodology used in each dataset determines which information is available as not all datasets contain information about unemployed and whether they receive benefits or not. Similarly, not all datasets contain information on individuals who are inactive. The specific way of reporting on these states in the datasets is explained in more detail in the next section.
- In some countries, the information contained in the dataset in relation to the states is richer. When this is the case (most notably, in the UK), additional states for alphabet B have been used. This allows to include information that can be potentially useful, whilst ensuring comparability across countries.

Table 4: Availability of alphabet B states and adaptation to national datasets

Labour market states	**ES**	**DE**	**UK**	
State A	13-14	13-14	State A	14
			State B	13
State B	11-12	11-12	State C	12
			State D	11
State C	9-10	9-10	State E	10
			State F	9
State D	6-8	6-8	State G	8
			State H	4-7
State E	4-5	4-5		
State F	-Unemployed receiving benefits -Leave of absence	-Unemployed receiving benefits	Unemployed Maternity Leave	
State G	Unemployed (not receiving benefits) and inactivity	Unemployed (not receiving benefits)		
			Student Sick / Disabled Family care	

Source: Authors

Data characteristics and alphabets used in the four selected countries

France

Box 4. Main differences of French data in relation to other countries analysed

Period and duration: 2009-2014, thus only post-crisis years are covered.

- Information on pay available at beginning and end points in the data only, not throughout the period under observation (this also determined that analysis of careers in France is limited to changes in relation to labour market states, since only alphabet A can be used);
- No information on company size;
- No information on working time.

The FQP Survey (*Formation et Qualification Professionnelle*) is carried out every 10 years by the French national statistical institute (INSEE) and covers five years of labour market history of people aged 21 to 65 years old. The main goals of the survey are to allow analysis in the area of: efficiency of the educational system, relationship between education, training and employment, labour market insecurity, job mobility and occupational mobility and social mobility between generations. The FQP data were collected in 2015 but captured the labour market situation in 2014. The sample was drawn from the 2013 fiscal administrative database. In total, 26,861 individuals were considered as respondents in the survey.

In the FQP dataset, two categories of information allow to characterise career trajectories. First, the labour market situations in 2009 and 2014 (employment, education and training, unemployment, retired, leave, out of the labour market) include detailed characteristics in case the individual is employed. Second, it includes a calendar with the following states: type of spell (employment, unemployment, leave, retired, mixed, out of labour market) and type of employment contract in case an individual is employed. The 'mixed' state is an alternate of very short fixed-term contracts and very short unemployment spells. In total, 20 spells are described. The date of beginning and ending of each spell is provided, if known, as well as the number of spells for each individual. These different dimensions facilitate the elaboration of a monthly calendar for a period of 60 months.

The main advantage of this dataset is that it allows a longitudinal perspective over six years. In addition, it contains a monthly calendar of the main states in the labour market. The availability of a monthly calendar over such a long period is very scarce in France. Moreover, the size of the sample is quite large, allowing detailed analysis, and the richness of the questionnaire opens a wide range of analysis of job mobility and social mobility.

Nevertheless, there are three main limitations with regards to the FQP data. First, it is not possible to distinguish between part-time and full-time jobs. As a result, a simplified version of alphabet A was used. Second, earnings' data are only available for the starting and ending points, hence making it impossible to use alphabet B. Finally, regarding the unemployment state, the information related to unemployment benefits is not available.

The main task in the construction of the dataset was to build a monthly calendar from available information in the database.

Four groups of variables are used:

- Number of spells for each individual (half of the sample experienced one spell);
- Nature of spell: employment, unemployment, retired, leave, out of labour market, mixed (alternating short fixed-term job and short unemployment spell). At most 20 spells are described;
- Nature of employment spell: self-employed, permanent, fixed-term, apprenticeship or professionalisation contract, temporary/seasonal contract/trainee;
- Date of beginning and ending of each spell.

Alphabet A

As mentioned previously, the lack of data on working time, obliges to adapt alphabet A in France since it is not possible to combine the type of contract with working time (see Table 5).

Table 5: France: states in alphabet A

State	Comment
Open-ended contract	Unlimited duration
Fixed-term contract	Contains fixed-term contract and mixed state (alternating short fixed-term and short unemployment spell)
Self-employed	With or without employees
Temporary or seasonal contract (including training contracts) contract/trainee	Fixed duration
Leave of absence	Interruption in employment
Unemployment (receiving benefits)	Looking for a job and currently unemployed but receiving some form of unemployment protection benefit
Unemployment (not receiving benefits)	Looking for a job and currently unemployed but without any unemployment benefit support
Out of labour force / Inactive	Currently unemployed and not looking for a job

Source: Authors

Germany

Box 5. Main differences of Germany data in relation to other countries analysed

Period and duration: 2001-2008 and 2009-2016 periods covered in the data. In order to avoid the bias caused by sample attrition over a 15-year period, the sample has been split into two periods, which means they cannot be compared and therefore differences between the two periods are not exploited in the analysis.

- Almost no observations for individuals under the age of 25 (and very few for those below 36) after preparing the data for the analysis as only full careers considered;

- Very limited observations for individuals under the age of 35 after preparing the data for the analysis;

- Inactive people are not covered in the sample as only full careers considered;

- No information on leave of absence.

The dataset used for Germany is the German Socio-Economic Panel (GSOEP), more specifically version 33.1 released in January 2018. It contains a survey period of 1984-2016[1]. The GSOEP is a wide-ranging representative longitudinal study of private households, carried out by the German Institute for Economic Research, DIW Berlin. Around 30,000 respondents in nearly 11,000 German households are interviewed on an annual basis. The data provide information on all household members, consisting of Germans living in the old and new German states, foreigners, and recent immigrants to Germany. Although GSOEP covers a wide range of topics, it was primarily designed to advance studies on the life course and on well-being, which is measured by indicators on life satisfaction and income. For the purpose of life course research specifically the data are structured as a panel, meaning that respondents are repeatedly (on a yearly basis) interviewed. In order to supply adequate coverage on specific groups, subsamples on, for example, recent immigrants and high-income households have been introduced.

For the purpose of this analysis, data from 2001 until 2016 are taken and divided into two phases: the first one ranging from 2001 to 2008, and the second one from 2009 to 2016. Phase one includes 4,632 cases and phase two 3,730 cases with complete information on alphabet B. For the whole period of 16 years, 1,898 respondents with balanced and complete information on the detailed employment status are included. Because the sample consists of few cases with balanced information for all 16 years, the cluster analysis was conducted on the two phases separately.

[1] Detailed information on variables of GSOEP and its subsamples can be found on https://data.soep.de/.

Alphabet A

In the case of Germany, there is no information about internships and leave of absence (see Table 6). However, in the case of unemployment, the GSOEP allows to identify two unemployment situations based on whether the individual benefits from an unemployment protection scheme.

Table 6: Germany: states in alphabet A

State	Comment
Full-time open-ended	
Part-time open-ended	
Full-time fixed-term	
Part-time fixed-term	
Self-employment	
Unemployed (receiving unemployment protection)	
Unemployed (without unemployment protection)	

Source: Authors

In the case of Germany, all those individuals who at any point of the period covered by the data have uninformed periods, have not been considered in the analysis. In other words, only individuals with full careers have been included. Compared to the case of Spain, where it is assumed that uninformed periods between two informed periods corresponds to either inactivity or unemployment without receiving benefits, in the case of survey data it is not possible to assume this, as these states are already considered in the questionnaire.

Alphabet B

For the operationalisation of alphabet B, information on employment status is differentiated into the categories of permanent contract, self-employed, temporary work, apprenticeship, unemployment receiving benefits and unemployment with no benefits. Information on internship and leave of absence is not included into the operationalisation, since the GSOEP data do not cover that information. Moreover, the prepared data has no information on inactivity, meaning that State G for alphabet B in Germany covers only unemployed not receiving benefits (see Table 7).

It is important to note in the case of Germany only those individuals for which there is complete information over the period considered have been included in the sample. Moreover, because of the sample effect, the age groups 16-25 and 25-35 were significantly smaller in the 2001 to 2008 period compared to the one covering the years 2009 to 2016. This explains why the presence of these age groups in the sample being extremely low compared to that in other countries.

A variable on professional categories has been generated which contains values for upper tier (Higher Managerial and Professional Workers, Lower Managerial and Professional Workers), mid-tier (Routine Clerical Work, Routine Service and Sales Work, Small Self-Employed with Employees, Small Self-Employed without Employees, Self-Employed Farmers) and lower tier (Manual Supervisors, Skilled Manual Workers, Semi- and Unskilled Manual Workers, Agricultural Labour) workers. The work type variable includes full-time, part-time and marginal part time, a job model in Germany with monthly earnings up to €450 and no social insurance contributions. In order to group respondents by their wage, a variable containing the net hourly wage in relation to the sample's median of all years considered has been generated. It distinguishes three pay brackets for respondents with wages: first, <66% of the median; second, 66% to 120% of the median; and third, for income above 120% of median hourly net wage. The calculation of the median is based on the weighted monthly net income, recalculated to the hourly net wage, based on hours by contract.

For a full summary of how scores were calculated in alphabet B for Germany see Table 7 and for the range of scores in each state, see Table 8.

Table 7: Scores for ranking employment states in alphabet B: Germany

Variable	Score for alphabet
Contract	Permanent = 4 Fixed term = 2 Temporary = 1
Hours	100%-85% of full-time = 3 84%-35% of full-time=2 <35% = 1
Occupation	Professional and manager = 3 Intermediate = 2 Low or partly skilled = 1
Earnings	< 66% of the overall mean of medians = 1 66% < 120% of the overall mean of medians = 2 >120% of the overall mean of medians = 4

Source: Authors

Table 8: Germany: States in alphabet B

State	Score for alphabet B
State A	13-14
State B	11-12
State C	9-10
State D	6-8
State E	4-5
State F	Unemployed receiving benefits
State G	Unemployed (not receiving benefits)

Source: Authors

Spain

Box 6. Main differences of Spain data in relation to other countries analysed

Period and duration: the data analysed covers the period 2000-2016, making Spain the only of the four countries under analysis where both pre-crisis and post-crisis periods can be compared, as the sample is the same throughout unlike in Germany. The periods are split into two groups in the analysis: 2000-2008 and 2009-2016 so as to make comparisons across countries by period.

- As the only administrative data source of the countries under study, it includes a larger number of 'uninformed' periods in individual careers where there are gaps in information on specific individuals;

- Uninformed periods are treated as a residual variable that corresponds to two possible states: unemployed not receiving benefits or inactive. There is not enough information in the sample to distinguish between these two states;

- Disaggregated information on internship/training contracts available (not the case in the other three countries).

The dataset used for Spain is the Continuous Sample of Working Lives (Muestra Continua de Vidas Laborales, MCVL), a random sample of around 1.2 million people (around 4% of the reference population). It contains administrative data on working lives, which provide the basis for this sample taken from Spanish social security records and comprises anonymised microdata with detailed information on individuals. The sample is extracted in 2016, meaning that it can include people having a relationship with the Spanish social security during that year. For each individual in the sample, the information goes back as the administration has information, starting in the 1960's. The period considered is 2000-2016. This database is unique in the richness of information it includes on labour market histories and personal characteristics such as nationality, date and country of birth, gender and place of residence when the individual first entered the social security system, along with additional information about the composition of the household and labour market variables (type of contract, sector, occupation, earnings and characteristics of company where they were employed).

It is very important for the interpretation of the Spanish results to know that careers have been built retrospectively. This means that 2016 is taken as reference year including those who are in the labour market, and then careers obtained moving back until 2000. This implies that some of the individuals that are present in the labour market in 2016 weren't yet active in the labour market. As the number of newcomers to the labour market in the early 2000s was particularly high in Spain (due to intense inward migration and the incorporation of young workers into the labour market), this explains comparatively higher inactivity rates in Spain at the beginning of the period considered, that declines gradually, even during the economic crisis. A trend that can be appreciated in different parts of the analysis, and more graphically in the state-flow diagrams, is the fact that the inactive/unemployed segment of the sample slowly enters the labour market, despite the simultaneous job destruction process occurring in the country. In that sense, the sample shows a certain degree of "survival bias", where those currently active are represented in detriment of those affected by the crisis during the previous decade. This explains why there were less people that had informed careers over the seven-year pre-crisis period and even less that had a standard one. As people with uninformed periods are likely to fall under non-standard careers, this explains the relatively low percentage of standard careers in Spain in our analysis.

The main advantages of the dataset include its large size, the continuity of information on working lives (allowing both static and longitudinal study) and the inclusion of fiscal data. However, the MCVL also presents some limitations that are important to bear in mind when interpreting the results. First, there is a limited number of variables related to socio-demographic characteristics. Second, there are a lot of missing values, repeated records, inconsistencies and more generally, problems of debugging. Third, some groups in the labour market are not included: civil servants, workers who are not employed and do not receive pensions; unemployed who do not receive unemployment protection; people who leave the labour market. Finally, some of the variables, including the level of education, are not fully updated.

Compared to the rest of countries, where survey data has been used, the administrative data used for Spain does not allow to capture inactivity or those unemployed not receiving unemployment benefits. As the MCVL is based on administrative social security records, only those individuals maintaining a relationship with the Social Security system are captured. This includes those employed, but also the unemployed that received some unemployment protection benefits. This means that there is no information on inactivity. However, it is possible to approach inactivity through the uninformed period between two informed periods. In other words, if over the period considered an individual has been employed, say two years, then has no relationship with the social security for six months and then is back to employment, it is known that the six month period is due to either: a) inactivity, because the individual has left the labour market temporarily; or b) unemployment, but the individual didn't receive any benefit and therefore had no relationship with social security. That's why "Inactive and unemployed without benefits" are grouped under the same heading.

Alphabet A

In order to construct alphabet A, the variables and states shown in Table 8 have been considered.

Table 8: Spain: states in alphabet A

State	Comment
Full-time open-ended	
Full-time fixed-term	
Part-time open-ended	
Part-time fixed-term	

State	Comment
Self-employed	
Internship	
Unemployment (receiving benefits)	
Leave of absence	
Uninformed (Unemployment without benefits + inactivity)	This category is used to capture those who are unemployed but do not receive unemployment benefits together with those inactive, not being possible to distinguish the two categories

Source: Authors

Alphabet B

All four variables considered in order to construct alphabet B were present in the case of Spain and, given its administrative nature, are considered to be very reliable. The variables were categorised following the general indications for alphabet B indicated earlier. In the case of earnings data, the data for Spain provides social security contributions as a proxy for earnings (this may underestimate earnings for the higher incomes but it provides a good approximation to the earnings' structure in the economy).

For a full summary of how scores were calculated in alphabet B for Spain see Table 9 and for the range of scores in each state, see Table 10.

Table 9: Scores for ranking employment states in alphabet B: Spain

Variable	Score for alphabet
Contract	Permanent = 4 Fixed term = 2 Temporary = 1
Hours	100%-85% of full-time = 3 84%-35% of full-time=2 <35% = 1
Occupation	Professional and manager = 3 Intermediate = 2 Low or partly skilled = 1
Earnings	< 66% of the overall mean of medians = 1 66% < 120% of the overall mean of medians = 2 >120% of the overall mean of medians = 3

Source: Authors

Table 10: Spain: States in alphabet B

State	Score for alphabet B
State A	13-14
State B	11-12
State C	9-10
State D	6-8
State E	4-5
State F	-Unemployed receiving benefits -Leave of absence
State G	Unemployed (not receiving benefits) + Inactivity

Source: Authors

United Kingdom

Box 7. Main differences of UK data in relation to other countries analysed

Period and duration: only pre-crisis, 2001-2008.

- Disaggregated information available and used on more detailed labour market states, including: student; sick/disabled and different forms of leave – maternity, family care;
- No meaningful information on foreign-born individuals in the sample;
- No wage information, meaning a proxy on financial security was used in its place.

The dataset used in the analysis of the UK labour market is the British Household Panel Survey (BHPS), a long-term panel of individuals nested within households. It has been collected since 1991 and ended in 2008/2009, when it was absorbed into a new panel called the UK Household Longitudinal Survey (UKHLS). The questionnaire is radically different in the new survey and many of the variables collected in the BHPS are not available in the UKHLS thus making it difficult to integrate the two surveys for this analysis.

The BHPS data used for the analysis in this report date from 2002 to 2008. The sample includes adults aged 16 or above. There are several employment status states that can be assessed during the interview: employed, self-employed, maternity leave, unemployed, family care, full-time education, retired, permanently sick/disabled, government training scheme and other. Variables for working hours, contractual status (permanent versus temporary contract), occupation and financial situation are also available. The study does not use the income data as there are many missing values and these variables are notoriously unreliable. Data entries without labour market activity throughout the seven-year period are deleted. The same applies to those who are permanently classed as sick/disabled, family care or full-time students. Also, those who have a labour market status of 'other' or missing are dropped from the analysis. Finally, anyone who recorded a retired status at any point in the sequence is also deleted.

Alphabet A

In the case of alphabet A, it was not possible to differentiate between unemployed receiving and not receiving unemployment benefits (see Table 11). Moreover, in the case of inactivity, a distinction was made between different states (full-time student, family care and sick / disabled).

Table 11: UK: states in alphabet A

State	Comment
Full-time permanent	
Full-time temporary	
Part-time permanent	
Part-time temporary	
Self-employed	

Unemployed	Including those who receive and don't receive unemployment benefits
Maternity leave (leave of absence)	
Full-time student (Inactive)	
Family care (Inactive)	
Sick-Disabled (leave of absence)	

Source: Authors

Alphabet B

Alphabet B was based on four variables: type of relation to the labour market; working hours (full-time/part-time); professional category; and financial situation. The variables are recoded into three or four categories, ranked and given a score following the indications provided in Figure 12. The earnings variable was proxied through the financial situation: it is important to bear in mind that respondents will tend to under-estimate their financial position, therefore introducing a bias in the responses towards the lower categories in this variable.

Every person, at any given point in time, has a total score resulting from the addition of the combination of the four ranked variables in Table 12. The final score is lastly transformed from integer to ordinal, into the alphabet. This alphabet is enhanced by adding additional categories based on the extra employment status variables available in the BHPS. This results in a sample of 4,973 individuals for analysis for alphabet B.

Table 12: Scores for ranking employment states in alphabet B: UK

Variable	Score for alphabet
Contract	Permanent = 4 Fixed term = 2 Temporary = 1
Hours	30+ = 3 16-29 = 2 1-15 = 1
Occupation	Professional and manager = 3 Intermediate = 2 Low or partly skilled = 1
Financial situation	Living comfortably = 4 Doing alright = 2 Just about getting by or worse =1

Source: Authors

Alphabet B states in the UK have been defined in a slightly different way in order to better capture differences in the labour market (See Table 13). In this way, more states for employed situations were considered, but with a direct correspondence between the general set up for alphabet B and the UK adaptation.

Table 13: UK: states in alphabet B

State	Score for alphabet
State A	14
State B	13
State C	12

State D	11
State E	10
State F	9
State G	8
State H	4-7
Unemployed Maternity Leave Student Sick / Disabled Family care	0*

*Note: *all part of inactive*

Source: Authors

The European Foundation for the Improvement of Living and Working Conditions (Eurofound) is a tripartite European Union Agency established in 1975. Its role is to provide knowledge in the area of social, employment and work-related policies according to Regulation (EU) 2019/127.