



*Demographic Research* a free, expedited, online journal  
of peer-reviewed research and commentary  
in the population sciences published by the  
Max Planck Institute for Demographic Research  
Konrad-Zuse Str. 1, D-18057 Rostock · GERMANY  
[www.demographic-research.org](http://www.demographic-research.org)

---

***DEMOGRAPHIC RESEARCH***

VOLUME 9, ARTICLE 7, PAGES 119-162

PUBLISHED 17 OCTOBER 2003

[www.demographic-research.org/Volumes/Vol9/7/](http://www.demographic-research.org/Volumes/Vol9/7/)

DOI: 10.4054/DemRes.2003.9.7

*Descriptive Findings*

**Employment status mobility from a life-  
cycle perspective: A sequence analysis of  
work-histories in the BHPS**

**Miguel A. Malo**

**Fernando Muñoz-Bullón**

© 2003 Max-Planck-Gesellschaft.

## Table of Contents

1	Introduction	120
2	Empirical analysis of life cycle data	121
2.1	Aggregate analysis	121
2.1.1	Number of spells (total mobility)	125
2.1.2	Gross mobility measure	127
2.2	Descriptive sequence analysis	130
3	Estimation of the differences along the life cycle: Sequence analysis	136
3.1	Alternative methods	136
3.2	Background: The basics of Optimal Matching Analysis (OMA)	137
3.3	Optimal Matching Parameters: Substitution costs	140
3.4	Optimal Matching Analysis results	142
3.5	OLS regressions on sequence dissimilarities	144
4	Conclusions	149
5	Acknowledgements	151
	Notes	152
	References	156
	Appendix A	159
	Appendix B	160

*Descriptive Findings*

**Employment status mobility from a life-cycle perspective:  
A sequence analysis of work-histories in the BHPS**

**Miguel A. Malo<sup>1</sup>**

**Fernando Muñoz–Bullón<sup>2</sup>**

**Abstract**

In this paper we apply optimal matching techniques to individual work-histories in the British Household Panel Survey (BHPS), with a two-fold objective. First, to explore the usefulness of this sequence-oriented approach to analyze work-histories. Second, to analyze the impact of involuntary job separations on life courses. The study covers the whole range of employment statuses, including unemployment and inactivity periods, from the first job held to the year 1993. Our main findings are the following: (i) mobility in employment status has increased along the twentieth century; (ii) it has become more similar between men and women; (iii) birth cohorts in the second half of the century have especially been affected by involuntary job separations; (iv) in general, involuntary job separations provoke employment status sequences which substantially differ from the typical sequence in each cohort.

---

<sup>1</sup> Departamento de Economía e Historia Económica, Edificio FES. Universidad de Salamanca. 37007- Salamanca. Tel: +34 923 29 46 40 (ext. 3512). Fax: +34 923 29 46 86. E-mail: malo@usal.es

<sup>2</sup> Sección de Organización de Empresas. Universidad Carlos III de Madrid. C/ Madrid, 126. 28903-Getafe (Madrid). Tel: +34 91 624 58 42. E- mail: fmunoz@emp.uc3m.es

## 1. Introduction

Event history analysis constitutes one of the principal statistical techniques in demography. It focuses on the time-to-event as the dependent variable, and allows researchers to study the transition rate of moving from one state (e.g., unemployment) to another (e.g., employment). However, this tool does not consider the whole sequence of multiple transitions between states simultaneously. Thus, it is less able to answer questions about the structure and composition of “trajectories” across labour market states. In contrast, sequence analysis is useful because it provides a tool to identify clusters over the life cycle, which is not equally straightforward in duration analysis. The order of states is not explicitly taken into account in event history models. Given that employment states appear in a certain order, this limitation is important for the analysis of the life cycle. Therefore, when we want to analyze sequences taken as wholes, traditional event-history approaches are not well suited to the analysis of sequential data.

We take these limitations as an invitation to do further research on the analysis of mobility from innovative approaches. In contrast to stochastic models of specific transition probabilities, sequence analysis is useful for studying employment status sequences taken as wholes. The objective is to empirically identify a typology of sequences, by taking into account the full complexity of sequence data. This is precisely what we do here. The aim of this paper is to apply sequence analysis techniques in the British Household Panel Survey (BHPS) to explore employment status mobility as embedded in the individuals’ life course. As we will see, our analysis offers complementary information to event history models and provides a more convenient method in order to get an overview about the pattern of sequences of labour market states.

A second motivation for this article is that, apart from the studies on the dynamics of employment statuses for females —where the focus is placed on the discontinuity of life-time work experience as an important source of the wage gap between males and females (See note 1)— research dealing with the historical changes of employment status mobility for men in Britain is much scarcer. However, the importance of involuntary job separations for both men and women must be underlined for European countries, where more mobile employment histories might appear to be chiefly caused by rising labour market flexibility. Given that redundancies —and, in general, all involuntary job separations— create long-term effects on subsequent workers’ labour market experience and their earnings’ profiles (See note 2), our analysis will explicitly consider whether or not individuals have ever suffered an involuntary job separation along their work-life history. In this respect, we make two principal contributions. First, we show how sequence analysis techniques facilitate a more comprehensive

understanding of those mobility patterns, which distinguishes our paper from earlier British research about job mobility (See note 3). Our second contribution is to provide new information about historical employment mobility changes in Britain, placing an special focus on the incidence and consequences of involuntary job separations. We demonstrate that involuntary job separations present a substantial impact on individuals' labour market mobility.

The popularity of optimal methods in the Social Sciences varies considerably by discipline. In Sociology, it has been mainly applied to the analysis of labour careers (see Abbot and Hrycak, 1990) and class careers (see Halpin and Chan, 1998). Both types of careers have been considered in Sociology as a finite set of temporarily ordered patterns. In general, sociologists have long argued that certain life course and occupational processes follow such scripts, and optimal alignment methods suit very well with this conceptual framework (See note 4). In Economics, use of these methods is extremely rare. Probably, the main reason is that sequence analysis is data driven, which does not conform well to the traditional approach in Economics. An empirical economic analysis requires that the empirical model can be derived from economic theory, and economic theory is not built to make predictions on sequences of events but on transitions between events. Life course is not usually understood as a sequence of states, but as the framework of some key transitions (from school to work, from temporary to permanent jobs, from work to unemployment, from activity to retirement, etc.).

The remaining part of the article is organized as follows. In the next section, we use two alternative measures of mobility — the spell, on the one hand, and the sum of all employment status changes, on the other— to investigate the relationship between sample cohorts' employment status mobility and involuntary job separations. In addition, a description of employment status sequences is undertaken with two purposes: first, to identify distinct patterns of life cycle sequences for the whole sample and its different birth cohorts; second, to compare the sub-sample of individuals who have ever suffered involuntary job separations with the sub-sample of those who have not. In the third section, Optimal Matching Analysis is applied and results obtained are presented. Finally, a conclusions section resumes the main insights of the paper.

## **2. Empirical analysis of life cycle data**

### **2.1. Aggregate analysis**

Our data are from the first three waves of the British Household Panel Survey (BHPS). The first wave was designed as a nationally representative sample of the population of Great Britain living in private households in the Autumn of 1991 (Northern part of

Scotland was not included). Approximately, 5,500 British private households (containing about 10,000 persons) were interviewed. These original sample respondents have been followed (even if they split off from their original households) and they, and their adult co-residents, interviewed at approximately one year intervals subsequently.

Information is recorded on labour market status at each interview, and for the period between 1 September a year before and the interview date. Thus, for respondents present at waves 1 to 3, we have a complete and detailed record of their labour market status from 1 September 1990 (or before: the start date of a job held at that date is known) to at least 1 September 1993. In addition, for our life course analysis, it is also necessary to have information on the respondent's entire career. In order to fill the gap since leaving full-time education to the start of the panel-derived labour market history, retrospective data were also collected in waves 2 and 3. In wave 2 a complete employment status history was collected, recording non-employed states in detail, and in wave 3 a complete job history was collected with detailed information on every job held (See Appendix A for documentation on the data sets from the BHPS used in this paper). Thus, we can construct a complete employment/labour market status history for nearly every individual in the survey from his/her first job to the year 1993 (See note 6).

The analysis reported in this paper uses a sub-sample consisting of all original sample members aged at least 34 years-old at 1-December-93 (so as to avoid right-censoring problems for very young individuals; see note 5). The sub-sample with non-missing information on the covariates used in the empirical analysis consists of 6,159 individuals (see Appendix A for sample statistics; See note 7). In principle, recall bias is a problem for our analysis. However, in practice, previous research attempting to assess the magnitude of recall effects in the BHPS has not found evidence of any significant bias. Indeed, it has been argued that much of the recall error can be described as random error, the exception being for short duration events —especially unemployment (See note 8). This can result in a biased and inaccurate account of cumulative experience, but need not be any worse than error inherent in data collected by panel methods. The BHPS has also attempted to minimize recall error by asking sample members to detail marital and fertility events (which tend to be well remembered) prior to their employment histories, thereby providing a chronological ordering of personal histories aiding the recall of employment events. This procedure has been shown to work well in other surveys. Hence we argue that the recall error in the BHPS labour histories is less of a problem than in most other retrospective data sets.

As the facts analysed in the article exhibit a close connection to individuals' life cycle, an explicit consideration of the different birth cohorts will allow a better understanding of the results. The definition of the different cohorts is as follows:

- First cohort: individuals who were born between 1906 and 1919.
- Second cohort: individuals who were born between 1920 and 1929.

- Third cohort: individuals who were born between 1930 and 1939.
- Fourth cohort: individuals who were born between 1940 and 1949.
- Fifth cohort: individuals who were born between 1950 and 1959.

Table 1 presents some relevant cohort characteristics. Most of individuals in the first two cohorts—and partially those in the third one—are above the mandatory retirement age. Thus, we are able to observe the complete life-cycle evolution of their employment status dynamics. On the contrary, life cycles must be considered as ‘right-censored’ in the remainder cohorts. The starting average year of the first spell offers an idea of the problems—or advantages—that each cohort must face in their eventual entry into the labour market. Whereas the first cohort starts their work-life histories at the beginning of the Great Depression, the second one does amidst the Second World War, the third one in the early fifties—i.e., while the economy was recovering from the previous recession period—the fourth one in the early sixties, and finally, the last cohort’s first spell is fairly close to the first oil shock. In order to appreciate how certain exogenous events might have affected the cohorts’ labour market evolution, Table 1 also reports their average age at the first and second oil shocks. The first two cohorts must not have been substantially affected by these shocks, whereas the remainder ones have presumably suffered the consequences of the oil crisis—especially the last two cohorts— either through redundancies, or through longer and more frequent unemployment spells, or both. (See note 9).

**Table 1:** *Birth cohort characteristics*

	Cohort 1 (1)	Cohort 2 (2)	Cohort 3 (3)	Cohort 4 (4)	Cohort 5 (5)
Age at 3 <sup>rd</sup> wave	74-87	64-73	54-63	44-53	34-43
Starting average year of 1 <sup>st</sup> spell	1929	1941	1951	1963	1972
Avg. age at starting year of 1 <sup>st</sup> spell	16	16	17	18	18
Avg. age at 1 <sup>st</sup> oil shock (1973)	60	48	38	28	18
Avg. age at 2 <sup>nd</sup> oil shock (1979)	66	54	44	34	24
Number of observations	905	1207	1124	1475	1508

Notes:

"Avg." means Average; (1) 1906-19; (2) 1920-29; (3) 1930-39; (4) 1940-49; (5) 1950-59.

Since our focus is placed on analysing whether job interruptions are associated with higher life course mobility, we have constructed two mobility measures in order to compare the behaviour of those sample members who have ever experienced an involuntary job separation with those sample members who have not (See notes 10 and 11).

The first mobility measure is the number of experienced spells —either employment, inactivity or unemployment spells. For instance, an individual with a sequence of labour market states as ‘Unemployment – Employment – Employment – National Service – Employment – Employment’ has a total of six spells and this mobility measure would then equal to six. We will address this measure as total mobility. The higher this number, the higher the labour market mobility the individuals will have been passing through. The reason why this mobility measure is appropriate for our analysis lies on the fact that individuals who exhibit a relatively low number of spells will have undergone a situation of almost permanent employment, inactivity or unemployment. Therefore, a low number of spells can be assimilated to a rather ‘static’ life course.

We have defined the second mobility measure as the number of changes observed in the individual’s labour market status sequence between successive spells with a *different* labour market status. Therefore, job-to-job changes are not considered by this mobility measure, because those are not changes in the labour market status of the individual. For instance, the individual with the sequence of the previous example (‘Unemployment – Employment – Employment – National Service – Employment – Employment’) experiments three changes. This mobility measure (which we address as gross mobility measure) would then equal three. These changes can be understood as changes in their attachment to the labour market. Table 2 shows the different degrees of labour market attachment (See note 12). The assigned numbers only indicate the *degree* of the attachment. For example, value 4 means that the attachment to the labour market in the corresponding spell is larger than for value 2, though, of course, not exactly twice as much. The individual of our example will have the sequence “455055” or “DEE0EE”. The gross mobility measure shows how many times the individual experiments a transition in his/her labour market statuses by comparing the attachment variable in two successive spells (See note 13). Its numerical value refers to changes, though not to the direction of the movement (i.e., whether an increase or a reduction in labour market attachment is observed).

**Table 2:** *Coding of labour market status*

Coding	Description
0 (O)	National/War Service
1 (A)	Out of labour force: retired; long-term sick; disabled
2 (B)	Out of labour force: maternity leave; family care
3 (C)	Full time student
4 (D)	Unemployment; government training scheme (“hidden” unemployment)
5 (E)	Employment: self-employed, full-time paid employment; part-time paid employment

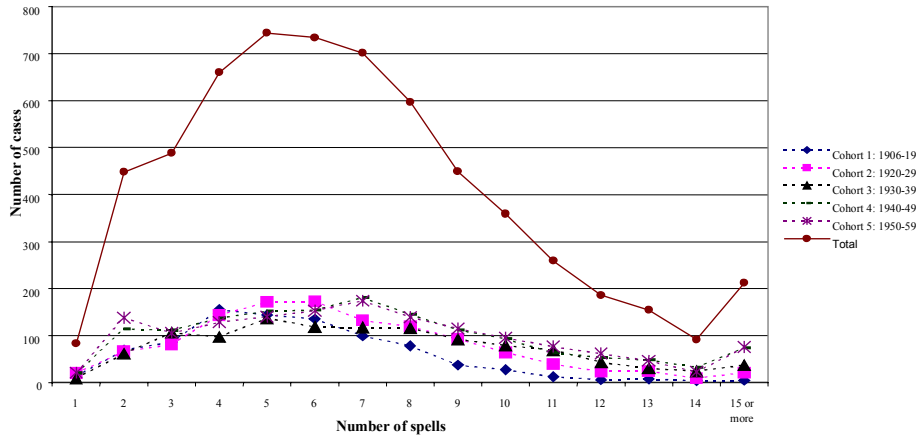


### **2.1.1. Number of spells (total mobility)**

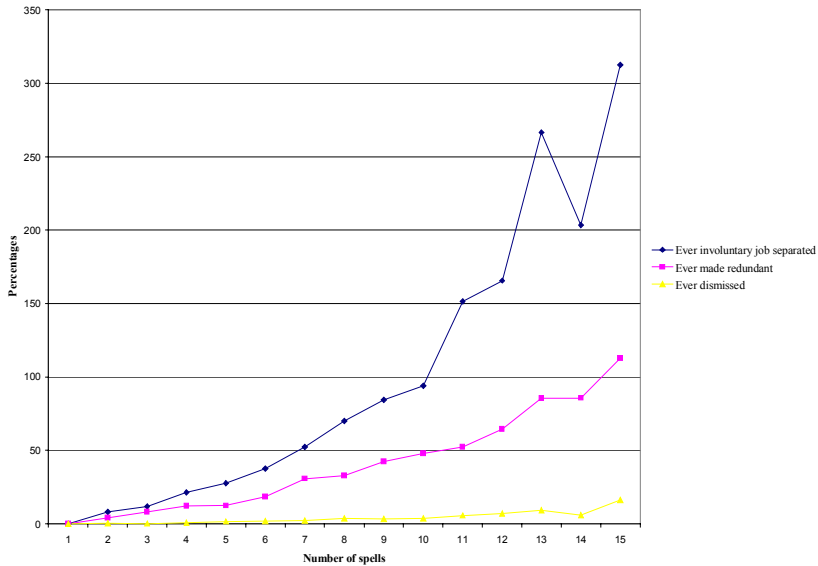
In this section we discuss the main descriptive features of the aggregate number of all types of spells (employment, inactivity, etc.) in order to have an approach to total mobility. Figure 1 shows the frequencies of the number of spells by cohorts and for the total sample. The maximum number of spells is forty three. Most individuals have five spells, although four, six and seven spells follow closely. Observing individuals with more than fifteen spells is highly unlikely. In addition, individuals of different cohorts behave differently. Whereas the two youngest cohorts' mode is larger than the eldest's one, the third cohort follows a mixed pattern: it presents a relatively high number of spells, although its evolution is closer to the fourth and the fifth cohorts. At first sight, therefore, the youngest cohorts reveal a more dynamic employment status history than the eldest ones.

We would like to point out that this mobility measure can be affected by a sort of 'right-censoring', because only those individuals who are already retired will have completed life courses. Up to our knowledge, the problem of censoring has not yet received any answer in sequence analysis. In this article, our way of dealing with this issue is to consider that the right-censoring of life courses is related to the birth cohort, and to assume that all individuals belonging to the same cohort are affected by a similar right-censoring problem. In any case, even when we consider the fact that younger cohorts have had less time to experiment mobility, Figure 1 shows that the fourth and fifth cohorts yield higher frequencies than the remainder cohorts for 7 or more spells.

Do ever involuntarily job separated individuals exhibit more spells than those never involuntarily job separated? Figure 2 shows the proportion between the former and the latter groups of sample members by number of spells. It also plots the proportion between sample members ever made redundant (dismissed) and those never made redundant (dismissed). Note that individuals who have ever been involuntarily job separated account for less than 100% of those never involuntarily job separated when the number of spells is below 10. While for 10 spells onwards, a proportion over 100% is always encountered. This increasing tendency also appears in the graph showing the proportion between individuals ever made redundant and those never made redundant. Finally, the odds between those ever dismissed and those who have not is almost constant regardless the number of spells.



**Figure 1:** *Frequencies of number of spells by cohort*



**Figure 2:** *Frequencies of number of spells: odds between those ever involuntary job separated (made redundant; dismissed) and those never involuntarily separated (made redundant; dismissed).*

In order to meaningfully interpret our first mobility measure for the different cohorts, we have constructed, for each of these, the ratio between the number of spells that ever involuntarily job separated individuals have and that of non-separated individuals. Then, the resulting numerical values for the first and the last pair of cohorts are divided by the value obtained for the third cohort. The fact of comparing ever and never job separated individuals inside each cohort helps us to overcome the problem of right-censoring, since individuals in the same cohort are affected on average by the same censoring problem. Thus, the result of this comparison exercise will not be biased. Table 3 summarises the results of this comparison to the third (reference) cohort for the three groups of individuals, that is: the comparison across cohorts of the spell ratios between ever and never involuntarily job separated individuals (first column); and results of the comparison across cohorts of the spell ratios between ever and never made redundant individuals (second column), and between ever and never dismissed sample members (third column). As can be observed, an increasing trend is present in the three columns as we advance from the eldest cohorts towards the youngest ones.

**Table 3:** *Relative mobility and number of spells: odds ratios by birth cohort*

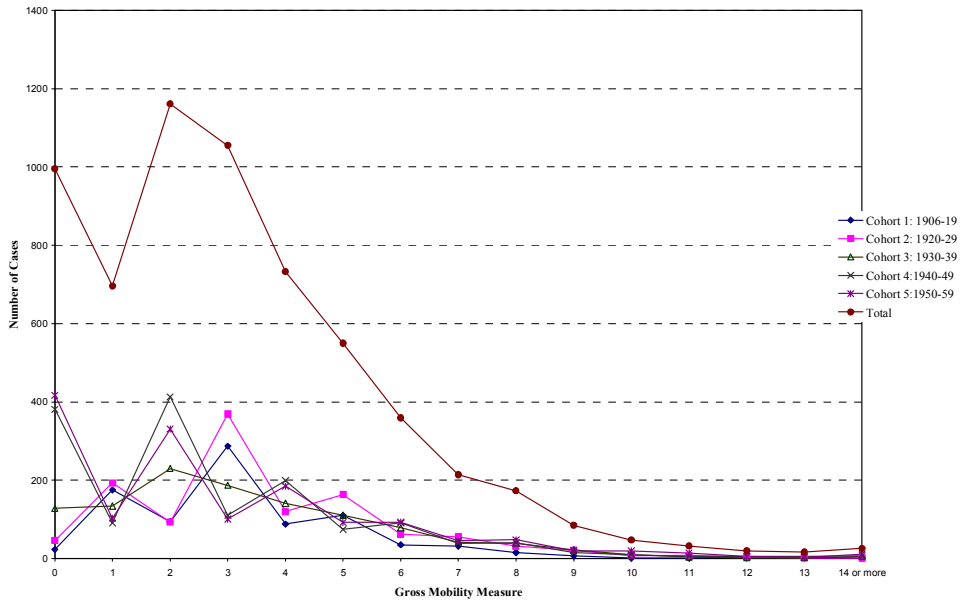
	Ever involuntarily job separated (1)	Ever made redundant (2)	Ever dismissed (3)
Cohort1 versus Cohort3	0.4	0.4	0.9
Cohort2 versus Cohort3	0.7	0.6	0.7
Cohort4 versus Cohort3	0.8	0.9	1.1
Cohort5 versus Cohort3	0.9	0.9	1.6

Notes:

- (1) Ratio between the number of spells of those ever involuntarily job separated and the number of spells of those never involuntarily job separated.
- (2) Ratio between the number of spells of those ever made redundant and the number of spells of those never made redundant.
- (3) Ratio between the number of spells of those ever dismissed and the number of spells of those never dismissed

### 2.1.2. Gross mobility measure

We now perform a similar analysis with the second mobility measure defined above (the so-called gross mobility measure). As we explained before, gross mobility is defined as the number of changes observed in the individual's labour market attachment across successive spells. Figure 3 plots the frequencies of this measure by cohorts and for the total sample.



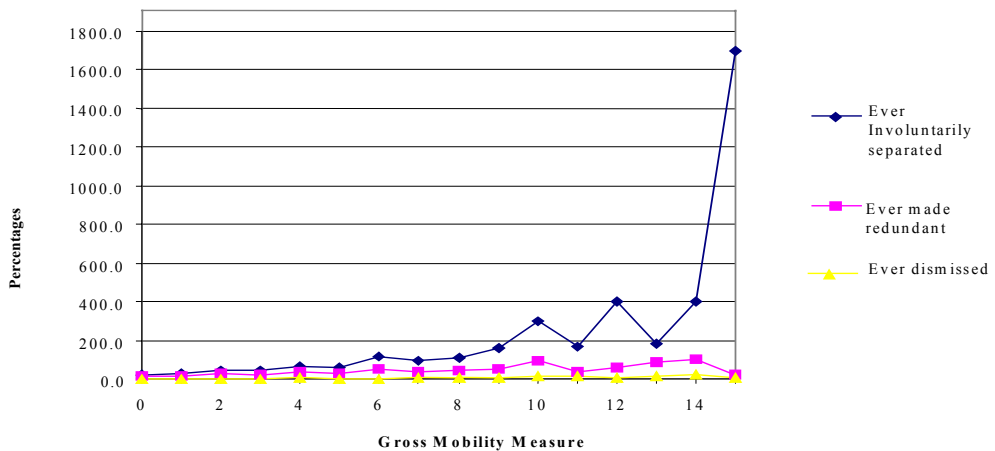
**Figure 3:** *Frequencies of gross mobility by cohorts*

This figure reveals a similar pattern to the one presented in Figure 1, although this time with a higher concentration on lower values, such as three and four changes. There are very few cases beyond ten changes. In addition, the cohorts show a different pattern up to five changes. From then on, all of them display very similar shapes. Finally, they attain their maxima at different peaks. The first two cohorts reach their maxima at three changes; the third cohort attains its maximum at two changes; finally, the fifth and fourth cohorts exhibit two maxima at zero and two changes, respectively (See note 14).

We may now wonder whether or not ever-involuntarily-job-separated individuals exhibit a higher number of transitions across labour market states than their non-separated counterparts. Two possible patterns can be obtained. On the one hand, it may be the case that mobility is lower among separated individuals. This might occur, for instance, when an initial separation leads the individual to a continuous situation of inactivity or unemployment. On the contrary, separated individuals may be able to go on active in the labour market, though perhaps only through sporadic work interspersed with frequent unemployment periods. In this latter case, following the initial separation,

workers will exhibit a larger number of transitions from employment to unemployment and vice-versa. That is, involuntary job separations will then be exerting a significant positive impact on job mobility.

Figure 4 shows the proportion of ever-involuntarily-job-separated individuals over the remainder of the sample by number of life cycle employment status transitions. It is evident how the former account for a higher proportion of the latter as the gross mobility measure increases. Though significantly weaker, this trend also appears when the proportion considers those individuals ever made redundant and those who have not. Finally, the odds between those who have ever been dismissed and those who have not is fairly constant independently of the number of transitions.



**Figure 4:** Gross mobility measure: odds between those ever involuntary job separated (made redundant; dismissed) and those never involuntarily job separated (made redundant; dismissed).

Table 4 shows the dynamics underlying the relationship between the gross mobility measure and the different cohorts. We have calculated, for each cohort, the ratio between the number of labour market transitions that ever-involuntarily-job-separated individuals present and that of the remainder sample members. Then, the resulting numerical values for the first and the last pair of cohorts are divided by the value

obtained for the third cohort. Table 4 summarises the results of this comparison to the third column (taken as a reference) for the three groups of individuals, that is: the comparison of the transition ratios between ever and never involuntarily job separated individuals (first column), and the results of the comparison across cohorts of the transition ratios between those ever and those never made redundant (second column), and between those ever and those never dismissed (third column). In general, an increasing trend is observed in the numerical values shown in the three columns as we advance downwards, in a similar fashion to the analysis of previous Table 3.

**Table 4:** *Relative mobility and gross mobility measure: odds ratios by birth cohort*

	Ever involuntarily job separated (1)	Ever made redundant (2)	Ever dismissed (3)
Cohort1 versus Cohort3	0.4	0.4	1.0
Cohort2 versus Cohort3	0.7	0.6	0.7
Cohort4 versus Cohort3	0.8	0.9	1.1
Cohort5 versus Cohort3	1.0	1.1	1.5

Notes:

- (1) Ratio between gross mobility measure of those ever involuntarily job separated and gross mobility measure of those never involuntarily job separated
- (2) Ratio between gross mobility measure of those ever made redundant and gross mobility measure of those never made redundant
- (3) Ratio between gross mobility measure of those ever dismissed and gross mobility measure of those never dismissed

Let us make several considerations as a partial summary of this aggregate analysis. The descriptive findings indicate that mobility has increased for the youngest cohorts. In addition, special events in work-histories —such as involuntary job separations— are associated with higher mobility in the youngest cohorts, but not in the eldest ones. At first sight, therefore, involuntary job separations are more disruptive for the youngest members of our sample. Finally, involuntary job separations, redundancies (and, in a minor extent, dismissals) increase the youngest cohorts' transitions across employment statuses along their life cycle. In any case, the proportion of the number of transitions accounted for by separated individuals does not rise as much as the proportion of the number of spells as we advance towards the youngest birth cohorts.

## 2.2. Descriptive sequence analysis

In this section we undertake a description of the sequence data constructed with the BHPS files, in order to generate typologies of labour market status sequences. The sequences of employment status spells have been constructed with the coding indicated

in previous Table 2. The symbol ‘.’ denotes that the last spell is the current spell at the time of the survey. For instance, the following string ‘EDEEA.’ represents a history with five spells—that is, the current spell is the fifth one—which started with an employment experience, after that the individual became unemployed, then she had two successive employment spells—with a job-to-job or promotion movement—and, finally, the individual is out of the labour force. The descriptive analysis of the sequences rapidly became very complex due to the fact that the maximum sequence length is forty three. In order to reduce this complexity, we have restricted the analysis to the first eight characters in each sequence.

In any case, standard distribution-based methods will not work simply as a consequence of the complexity of the analysis. In our sample, the probability that each two sample members can be represented by different sequences of states will almost always be close to 1. Moreover, given that sequences do not have a numerical meaning, we can not use the mean of the distribution of sequences as the mid-point of the distribution. For this reason, Table 5 shows both the median sequence and the mode sequence of the distribution of the full sample as well as that of the five cohorts for different sequence lengths. This procedure has allowed us to detect relevant regularities which are presented below (See note 15). We first present results for the whole sample, and, then, compare typologies of sequences by cohorts.

For the full sample, the most frequent sequence up to seven spells corresponds to full stability in employment by job-to-job movements. The presence of this pattern quickly decreases with the sequence length (for instance, while among two-spell sequences the pattern “EE” accounts for 45.1% of the observations, among five-spell sequences the pattern “EEEE” only accounts for 10.6%). As regards eight-spell sequences, the mode corresponds to exhibiting only two successive employment spells (with a frequency of 4.0%). We have checked that the pattern ‘EE.’ corresponds to sample members who have reached the moment of the survey in their second spell.

Results clearly differ by cohorts. In particular, they show that the patterns for the total sample are originated by the behaviour of the youngest cohorts. The youngest cohorts (third, fourth and fifth ones) show the two following main patterns: a history containing two spells (‘EE.’) and a history of successive jobs (‘EEEE\*’), where ‘\*’ means that successive spells have the same employment status as the last one. The former pattern constitutes a non-finished work-life experience since all of them fall below the mandatory retirement age (between 33 and 62 years-old).

Among two and three-spell sequences, all cohorts are fairly similar, although the youngest ones enjoy employment spells more frequently. For instance, the pattern ‘EE’ accounts for more than 54% of observations among the two youngest cohorts, while its frequency falls below 34% in the two eldest ones. The most frequent two-spell patterns are ‘EE’ (employment-employment) and ‘EB’ (employment-inactivity). We have analysed their distribution by gender. While there are only three cases of males in ‘EB’

sequences, 53.8 per cent of ‘EE’ sequences corresponds to males. Thus, the minor presence of the pattern ‘EB’ for the youngest cohorts might be due to the fact that women are showing work-life histories more similar to those of men.

As regards five and six-spell sequences, it is also possible to detect some life cycle patterns. On the one hand, there are two clear patterns in the first and second cohorts: a two-spell history (‘EA.’) and a four-spell history (‘E0EA.’). Both are non-complete work histories, but they are similar considering that the zero of the second history corresponds to National/War Service. Therefore, those two patterns are actually quite similar. If there was a recall bias affecting the answers of the eldest cohorts—which would then collapse successive employment spells into only one—the two eldest cohorts’ patterns ‘EA.’ and ‘E0EA.’ might be regarded as similar to the most frequent patterns of the youngest cohorts (‘EE.’ and ‘EEEE\*’). As explained in Section 2.1 above, previous research about recall bias using the BHPS has not found in particular this kind of bias (see note 16). Nonetheless, it is likely that the eldest cohorts forget shorter employment spells more easily, remembering the longest employment spells with higher probability (as Booth *et al.*, 1999 remark). In addition, since for the youngest cohorts these shorter employment spells must be closer in time than for the eldest ones, they would be more likely included in the pattern of successive employment spells. As a consequence, having only two employment spells (‘EE.’) can be regarded as a work-history fairly similar to the sequence ‘EA.’. However, the mode ‘EEEE\*’—typical of the youngest cohorts— may be suggesting a new work-life history pattern as opposed to the eldest cohort sequence ‘E0EA.’, the reason lying on the fact that it is highly unlikely that individuals in the eldest cohort have forgotten every spell except for one (See note 17).

**Table 5:** Medians and modes of labour market status sequences

FULL SAMPLE Sequence Length	MEDIAN		MODE	
	Sequence	% (Cum.)	Sequence	%
2	EC	48.1	EE	45.1
	ED	54.9		
3	EDD	49.1	EEE	25.0
	EDE	54.9		
4	EDEA	49.9	EEEE	15.6
5	EDEBB	50.0	EEEEEE	10.6
6	EDEBBE	50.0	EEEEEEE	7.4
7	EDEBBEB	50.0	EEEEEEEE	5.3
8	EDEBBEBE	50.0	EE.	4.0
			EEEEEEEE	3.4
<b>1st. COHORT</b>				
2	EA	39.9	EE	30.6
	EB	62.9	EB	23.0



3	EBB	47.6	EBE	15.3
	EBE	62.9	EEE	13.7
4	EBE	47.8	EBEB	6.6
	EBEA	52.2	E0EA	5.6
			EA	5.6
			EEEE	5.6
5	EBEA	49.4	EA	5.6
	EBEAA	51.3	EBEBE	4.0
			E0EA	3.8
			EEEE	3.0
6	EBEA	49.4	EA	5.6
	EBEAA	51.2	E0EA	3.8
			[EEEEEE]	[1.5]
7	EBEA	49.4	EA	5.6
	EBEAA	51.2	E0EA	3.8
			[EEEEEEE]	[0.8]
8	EBEA	49.4	EA	5.6
	EBEAA	51.2	E0EA	3.8
<b>2nd COHORT</b>				
2	EA	41.4	EE	33.5
	EB	62.0	EB	20.6
3	EBB	44.9	EBE	17.2
	EBE	66.2	E0E	16.6
			EEE	16.0
4	EBEA	48.9	EEEE	8.4
	EBEB	55.8	E0EE	8.1
			EBEB	6.9
			EBEE	6.1
5	EBEB	48.9	EEEE	4.8
	EBEBA	50.1	EBEBE	4.8
			E0EEE	4.4
			EA	4.2
6	EBEBAA	49.5	EA	4.2
	EBEBAB	50.1	E0EA	3.3
			[EEEEEE]	[2.9]
7	EBEBAB	49.6	EA	4.2
	EBEBABA	50.1	E0EA	3.3
			[EEEEEEE]	[1.9]
8	EBEBAB	49.6	EA	4.2
	EBEBABA	50.1	E0EA	3.3
			[EEEEEEE]	[1.5]
<b>3rd COHORT</b>				
2	EA	26.7	EE	43.1
	EB	50.5	EB	23.8
3	EBD	29.7	EBE	20.8
	EBE	50.5	EEE	20.6
			E0E	13.7
4	EBED	39.7	EEEE	11.4
	EBEE	50.5	EBEE	10.7
			E0EE	10.2
			EEBE	7.5
			EBEB	6.5

5	EBEED	44.4	EEEE	7.7
	EBEEE	50.5	E0EEE	6.2
6			EBEEE	6.1
			EBEBE	4.8
	EBEEEE	46.8	EEEEEE	5.6
	EBEEEE	50.5	E0EEEE	4.9
			EBEEEE	3.6
7			[EA]	[1.5]
	EBEEEEA	48.2	EEEEEEE	4.2
	EBEEEE	50.5	E0EEEE	3.7
			[EA]	[1.5]
8	EBEEEEEA	49.0	EE	2.8
	EBEEEE	50.5	EEEEEEE	2.7
			[EA]	[1.5]
<b>4th. COHORT</b>				
2	ED	44.1	EE	55.9
	EE	100	EB	23.1
3	EDE	44.1	EEE	33.5
	EE	51.2	EBE	20.0
4			[E0E]	[0.9]
	EDEE	44.1	EEEE	24.0
	EE	50.7	EBEE	11.8
5	EDEEE	44.1	EEEE	16.7
	EE	50.7	EBEEE	7.0
6			EE	6.6
	EDEEEE	44.1	EEEEEE	12.0
	EE	50.7	EE	6.6
7			EBEEEE	4.2
	EDEEEEE	44.1	EEEEEEE	8.6
	EE	50.7	EE	6.6
8			[EA]	[0.2]
	EDEEEEE	44.1	EE	6.6
	EE	50.7	EEEEEEE	5.7
<b>5th COHORT</b>				
2	ED	45.9	EE	54.1
	EE	100	EB	17.3
3	EDE	45.9	EEE	33.8
	EE0	53.7	EBE	13.4
4	EDEE	45.9	EEEE	22.3
	EE	53.3	EE	7.5
			EBEE	7.1
5	EDEEE	45.9	EEEE	15.9
	EE	53.2	EE	7.4
6	EDEEEE	45.9	EEEEEE	11.6
	EE	53.2	EE	7.4
7	EDEEEEE	45.9	EEEEEEE	8.4
	EE	53.2	EE	7.4
8	EDEEEEE	45.9	EE	7.4
	EE	53.2	EEEEEE	5.0
			E	

Notes:

"0": National/War Service; "A", "B", "C": Inactivity; "D": Unemployment; "E": Employment (See Table 2 for a more detailed description of those categories); '.' the last spell is the current at the time of the survey

An additional remarkable aspect in Table 5 is the movement of the median closer to the mode for the youngest cohorts. A plausible interpretation for this result lies on the fact that the distribution of sequences has changed, so that histories are more linked to the labour market (See note 18) and, in addition, sequences which are more homogeneous than before are being observed (See note 19).

**Table 6:** *Medians and modes of four-spell sequences by cohort and involuntary job separations*

Separation Status	Full Sample		1st COHORT		2nd COHORT	
	Median (%C)	Mode (%)	Median (%C)	Mode (%)	Median (%C)	Mode (%)
Ever Involuntarily Separated	EDED (45.7)	EEEE (21.9)	EBEE (49.7)	EEEE (9.3)	EBED (49.1)	E0EE (15.7)
	EDEE (51.0)	EBEE (6.5)	ECE0 (50.5)	E0EE (6.3)	EBEE (53.9)	EEEE (13.9)
Never Involuntarily Separated	ECDC (49.7)	EEEE (12.5)	EBB0 (48.8)	EBEB (6.8)	EBE0 (47.5)	EBEB (7.2)
	ECDE (50.1)	EBEE (9.0)	EBBA (50.9)	EEBE (6.4)	EBEA (51.7)	EBEE (6.7)
				EBEE (6.3)		

Separation Status	3rd COHORT		4th COHORT		5th COHORT	
	Median (%C)	Mode (%)	Median (%C)	Mode (%)	Median (%C)	Mode (%)
Ever Involuntarily Separated	EDED (49.6)	EEEE (16.3)	EEBD (43.7)	EEEE (32.3)	EDEE (49.5)	EEEE (26.9)
	EDEE (52.1)	E0EE (11.1)	EEBE (51.3)	EBEE (8.4)	EE00 (50.0)	EDEE (9.0)
Never Involuntarily Separated	EBED (41.9)	EBEE (12.6)	EDEE (46.2)	EEEE (19.5)	EDEE (43.9)	EEEE (19.8)
	EBEE (54.6)	E0EE (9.7)	EE00 (56.0)	EBEE (13.6)	EE00 (55.2)	EE00 (11.3)

Notes:  
 %C = Cumulative percentage.  
 % = Frequency

Finally, we introduce a distinction between individuals who have ever been involuntarily separated and individuals who have not in Table 6. This Table shows the medians and modes for the full sample and for each of the five cohorts. Only four-spell sequences are displayed, since patterns were clear enough for sequences with more spells. Sequences in both groups are rather similar in the fourth and fifth cohorts. However, they clearly differ from those in the eldest ones. Among the latter, the mode of those who have never been involuntarily job exhibits the pattern ‘EB’ in the first two spells which, as was explained above, corresponds to a typical female pattern.

As a summary, this descriptive analysis shows that the youngest cohorts are more mobile, that their histories are more linked to the labour market, and that they present more homogenous histories. The changes in womens’ sequence patterns may be explaining both their higher attachment to the labour market and the higher homogeneity in the sequences arising among the youngest cohorts. In addition, the type of job separation is found to be related to gender, given that involuntary job separations exert a relevant influence on men or, at least, on the typical ‘male’ histories.

### 3. Estimation of the differences along the life cycle: Sequence analysis

In this section we apply an *optimal matching analysis* (OMA) which draws from techniques used in DNA sequence matching (note 20). It was first introduced to the social sciences by Andrew Abbot (note 21), who has applied OMA to a range of sociological issues, such as the development of the welfare state and musicians' careers (note 22). The technique has also recently been utilized to study social mobility (note 23) and to examine other social and political processes (note 24).

#### 3.1. Alternative methods

Researchers with sequence data sets have at their disposal a variety of methods to be applied. Those methods can be classified in two types: step-by-step methods and whole sequence methods. Among the first group we encounter time series, Markovian and event history methods. Time series are applied when the interest is to find a simple stochastic generator that effectively fits entire sequences (which may involve auto-regression, moving averages or both; see note 25). Markov methods are applied in order to fit sequences of categories by estimating transition probabilities step by step (note 26). Finally, event history methods are used when the researcher's interest focuses on transitions from only one particular prior category and when the issue is time until transition (note 27).

The second group of tools —sequence methods— use the entire sequence as the unit of analysis and concentrate on questions of global similarities with a view to establishing sequence typologies. This methodology offers a powerful technique for analysing the full richness of sequence data without discarding the details of episode ordering, duration or transition. In particular, Abbot's optimal matching analysis — which we apply below in this section— has been used to cluster life courses in different groups. This purpose of classification has become the principal approach to sequence data analysis in the social sciences. Indeed, sequence analysis is increasingly becoming a useful method for examining social phenomena that occur over periods of time. Nonetheless, it also presents several problems, some of the most important must be underlined. First, at present, the measure theory of whole sequence data is rather preliminary. Second, those methods are largely heuristic. Third, in clustering it is sometimes complicated to identify the reasons why some individuals are assigned to a specific group. Fourth, it is unclear how much this assignment depends on the distance between states, which in some cases has to be assigned subjectively by the researcher.

For those reasons, instead of taking our OMA results as input to some additional descriptive analysis (such as is usually done with clustering algorithms), we will show in section 3.5 below which individual and job-related characteristics lie behind the estimated measure of distance resulting from our OMA application.

### 3.2. Background: The basics of Optimal Matching Analysis (OMA)

Let us briefly make a comment on the intuitive idea underlying this process of *optimal matching* (See note 28). We consider sequences of states which are elements of a finite space state, say  $\mathbf{Y}$ .  $\mathbf{S}$  denotes the set of all finite sequences over  $\mathbf{Y}$ , meaning that:

$$\text{if } a \in \mathbf{S} \text{ then } a = (a_1, \dots, a_n) \text{ with } a_1, \dots, a_n \in \mathbf{Y}$$

$n = |a|$  is the length of the sequence. We desire to compare two sequences  $a, b \in \mathbf{S}$ . The basic idea is to define a set of elementary operations which can be used sequentially to transform one sequence until it becomes equal to the other sequence.

Let  $\mathbf{Z}$  denote the set of basic operations and  $a(w)$  the sequence resulting from  $a$  by applying the operation  $w \in \mathbf{Z}$ . We consider three elementary operations:

- a) Insertion:  $a(i)$  denotes the sequence resulting from  $a \in \mathbf{S}$  by inserting one new element (a state from  $\mathbf{Y}$ ) into the sequence  $a$ .
- b) Deletion:  $a(d)$  denotes the sequence resulting from  $a$  by deleting one element from this sequence.
- c) Substitution:  $a(s)$  denotes the sequence resulting from  $a$  by changing one of its elements into another state.

We can think of sequentially applying elementary operations to a given sequence. Let  $a(w_1, w_2, \dots, w_k)$  denote the new sequence resulting from  $a$  by applying first the elementary operation  $w_1$ , then  $w_2$ , and so on until finally  $w_k$ . Then, given two sequences,  $a, b \in \mathbf{S}$ , we can ask for a sequence of elementary operations which transform  $a$  into  $b$ .

In general, there will be many such sequences of elementary operations which transform  $a$  into  $b$ . Now, the intuitive idea for developing a distance measure for sequences is to look for the shortest sequence of elementary operations which transform  $a$  into  $b$ . A slightly more general approach is to evaluate the elementary operations by introducing  $c(w)$  as the cost of applying the elementary operation  $w \in \mathbf{Z}$ . We will assume that  $c(w)$  is between 0 and infinite. The cost of applying a sequence of elementary operations will be denoted by:

$$c[w_1, w_2, \dots, w_k] = \sum_{i=1}^k c[w_i] \tag{1}$$

Setting  $c[] = 0$  for no operation, we can formally define:

$$d_Z(a, b) = \min\{c[w_1, w_2, \dots, w_k] \mid b = a[w_1, w_2, \dots, w_k], w_i \in Z, k \geq 0\} \tag{2}$$

to measure the distance between the sequences  $a$  and  $b$ . This measure is by definition non-negative, and  $d_Z(a, b) = 0$  only if  $a = b$  (See note 29).

“Optimal matching” without further qualification normally means referring to this distance measure based on  $Z = \{i, d, s\}$  resulting in a metric distance. Of course, the distance measure also depends on the definition of the cost functions  $c[w]$ . These cost functions can be arbitrarily defined (as we will see below) with respect to the intended applications. As a special case, one can set:

$$c(i) = c(d) = 1, \text{ and } c(s) = 2 \tag{3}$$

The distance between two sequences  $a$  and  $b$  is then simply the number of indel operations (insertions and deletions) which are necessary to transform one sequence into the other. Once the costs of those three basic operations are established, the algorithm evaluates all possible solutions for each pair of sequences and returns the cost of the most efficient transformation path as the “distance” between the sequences. Pairs of sequences with small “distances” are similar to one another, while pairs with larger “distances” are more distinct.

An example that will show how this optimal alignment method uses the set of operations to align and transform pairs of sequences is the following. Consider two hypothetical sequences in Table 7.

**Table 7:** OMA Example (I)

POSITION	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Sequence 1:	4	4	4	4	5	5	5	5	6	6	6	6	6	1	1	1
Sequence 2:	3	3	3	3	5	5	5	5	1	6	6	6	6	6	1	1

In the comparison of Sequences 1 and 2, we find discrepancies in elements 1-5, 9-11, and 14. We could use a transformation that entails a series of substitutions. This would involve, for instance, substituting the 3 3 3 3 in elements 1-4 of the second sequence for the 4 4 4 4 in the first sequence, along with other substitutions later in the sequence. The cost of each individual transformation (e.g., substituting a 3 for 4 in the second element) is set by a-priori established costs of the different transformation operations. If we imagine that the cost of substituting one element for another is the difference in their numeric values ( $w(a_i, a_j) = |a_i - a_j|$ ) then the minimum total cost of aligning the sequences is simply the sum of the individual operations; here, the “distance” between the sequences would equal:  $(4*|3-4|)+(|3-5|)+(|5-6|)+(|1-6|)+(|6-1|)=17$ .

The algorithm could also conceivably try a transformation strategy that involves insertions, deletions and substitutions. In this comparison, we could delete the 1 in element 14 of Sequence 1, shift the remaining 13 elements one position to the right and then insert a 3 in the first element (Table 8).

**Table 8:** OMA Example (II)

POSITION	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
Sequence 1:	4	4	4	4	5	5	5	5	6	6	6	6	6	1	1	1	(original)
Sequence 1a:	3	4	4	4	4	5	5	5	5	6	6	6	6	6	1	1	(after insertion and deletion)
Sequence 2:	3	3	3	3	3	5	5	5	5	1	6	6	6	6	1	1	

After the insertion and deletion, we have a situation where Sequence 1a and Sequence 2 are mismatched at only elements 2-5 and 10, thus requiring fewer substitutions. However, if the cost of using an insertion or deletion is high in part because we are adding essentially unknown elements to an established sequence, then using a combination of insertions, deletions, and substitutions is likely to produce a less cost-efficient transformation (See note 30). Let’s say that the cost of insertions or deletion (i.e., indel cost) in this example is 9. The distance between Sequences 1 and 2 using a combination of indels and substitutions would equal  $(2*9)+(4*|3-4|)+(|1-6|)=27$ . In this case, the most efficient transformation involves only substitutions, though there are instances where simple insertions or deletions can result in the most cost-efficient path of transforming two sequences of equal length.

### 3.3. Optimal matching parameters: Substitution costs

Insertion and deletion costs are always identical in the algorithm used in this paper — which is described in the Appendix— and called *indel cost* (See note 31). However, the setting of substitution costs between sequence elements —i.e., the different labour market states— involves, in general, serious consideration of theoretical questions. Optimal matching requires an explicit parameterization of substitutions, since substitution costs may affect distances. One can use a *default cost function* where indel cost equals 1 and substitution cost equals 2. In this case, the alignment of two sequences provides, then, the longest common subsequence (See note 32). This option is an interesting starting point because it makes a substitution equivalent to two indel operations —a deletion followed by an insertion; that is, both cost the same. However, we consider that, in our case, the initial basis for setting those costs should be related to the relative importance of changing the labour market states along the work history of the individual. Even though the data shows numerous moves across labour market states in the period of observation some labour market moves involve a substantial change in the degree of attachment to the labour market, when compared to other types of changes. That is, some pair of states seem to be fairly closely connected by career mobility than others. It seems necessary, therefore, to add mobility information to our measures of state resemblance or dissimilarity.

Doing that involved the use of two different “distance” matrices. For the first of those matrices, substitution costs is calculated as the absolute difference  $w(a_i, a_j) = |a_i - a_j|$ ; the underlying assumption for this is based on the fact that it is desirable that two states are more similar the less drastic the associated change between the correspondent labour market states. That is, not all substitutions really “cost” the same. The difference between a state of “Out of labour force” (code A in Table 2) and a situation of “Employment” (code E) is greater than the difference between “Out of labour force” (code A) and a state of “Full time student” (code C). This fact explains why we should differentiate the costs of substitution. Work-life histories involving status changes characterized by low “vertical” distances across occupations will then be closely linked by the matching algorithms.

The second approach derives substitution costs from the frequency of transitions in the given sequence data; with this option, mobility is therefore added to our measure of job resemblance by making use of data-based substitution costs. This approach allows us to use a distance matrix which is data-driven rather than completely *ad-hoc*. This issue is relevant in OMA applications such as ours, where there are no accepted measures of quantitative differences between states to guide the estimation of the substitution costs. In those cases, if researchers rely on their own theoretical assumption, then an *ad-hoc* cost function is likely to be employed. This procedure has been the target of criticisms by some authors (See Wu, 2000). But a way to circumvent



this particular weakness of the matching method is to use whatever empirical data the researcher has available. In our case, this involved producing a “distance” matrix based on mobility information by creating a matrix of transitions between the different values of the labour market statuses (as defined in Table 2). To explain this idea, let us assume a sample of  $i=1, \dots, n$  sequences:  $y_i=(y_{i1}, y_{i2}, \dots, y_{iT})$ , where the elements are valid statuses in the state space  $\mathbf{Y}=\{ 1, \dots, q \}$ . We can then define:

$$N_t(\mathbf{x})= \text{Number of sequences being in state } x \text{ at } t \tag{4}$$

$$N_{t,t'}(\mathbf{x},\mathbf{y})= \text{Number of sequences being in state } x \text{ at } t \text{ and in state } y \text{ at } t' \tag{5}$$

Those quantities can be used to construct the referred substitution cost matrix; the underlying intuition is that two statuses are less different when there are, in the dataset, more transitions from one status into the other. One way of using this information about transitions would be to use:

$$p(x, y) = \frac{\sum_{t=1}^{T-1} N_{t,t+1}(x, y)}{\sum_{t=1}^{T-1} N_t(x)} \tag{6}$$

in order to define substitution costs as follows:

$$w(a_i, b_j) = 2 - p(a_i, b_j) - p(b_j, a_i) \text{ if } a_i \neq b_j \tag{7}$$

$$w(a_i, b_j) = 0 \text{ if } a_i = b_j \tag{8}$$

That is, the information on substitution costs is implemented through a matrix of transitions which classifies every move in the individuals’ life-cycle experiences. Since this substitution cost matrix must be symmetric, we obtained such a matrix by adding corresponding elements (i.e., the  $i,j$ th and the  $j,i$ th for all  $i$  and  $j$ ) and replacing both the  $i,j$ th and the  $j,i$ th with this sum. Finally, since transition figures rise with mobility closeness rather than with distance (i.e., cost), we turned the matrix into a dissimilarity matrix by subtracting each element from a constant. Then,  $0 < w(a_i, b_j) < 2$ , and substitution cost directly reflects the cumulated transitions across statuses. If there are many transitions from  $a_i$  to  $b_j$  (both directions), the substitution cost will be low, and vice versa.

### 3.4. Optimal matching analysis results

Given a model of substitution costs, we can apply the optimal matching algorithm to our data. Pair-wise comparison of  $n$  sequences requires  $n(n-1)/2$  alignments. However, as Abbot and Forrest remark (1986), there is an inherent problem with analyses which involve procedures of pair-wise comparison: the processing time required rises with the square of, rather than in direct proportion to, the number of cases (see also Halpin and Chan, 1998). If  $n$  is big and if sequences are particularly long—as in our dataset—this results in strong constraints on the size of the sample we can deal with. We managed to circumvent this problem by taking into account the aim of our research: to analyse how much life cycle work-histories are disrupted by involuntary job separations. If we had calculated distances between all pair of sequences in our sample—apart from the computational burden involved—this would have offered no measure of dissimilarity between the two relevant sub-samples (ever involuntary job separated and the remainder of the sample). Thus, rather than calculating distances between all pairs of sequences, here we calculated the distance between each sub-sample's sequence and each of the median sequences in each cohort. Given that the deviation of the distribution of sequences from the mean is likely to be of considerable magnitude (see previous Section 2.2), we regard the median sequence in each cohort as the centre of the distribution, thereby providing an idea of the typical work-history in the cohort. In addition, this procedure allowed us to *substantially* reduce the computational intensity of the task, and to deal with an eventual 'right-censoring' bias affecting the length of sequences (given that we compare sequences inside each cohort, the censoring problem is mitigated, as we explained in previous sections).

Table 9 shows the first application of OMA to our data. The three measures of substitution costs defined in the previous section are used in order to optimally compare by gender all sequences in each sub-sample—individuals ever involuntarily separated (EINV1) and those never involuntarily separated (EINV0)—to the median sequence of the corresponding cohort. We have then calculated the mean individual distance in each sub-sample to their median sequence. Finally, the average individual dissimilarity is computed across individuals in each cohort. This final value is showed in Table 9 for each cohort and each of the distance matrices used—named 'Default', 'Absolute Value' or 'Data-Based'. For instance, according to the default substitution cost matrix, men in the first cohort who have never been involuntarily job separated have an average distance of 3.19 to the median men's sequence in this first cohort. Should we consider individuals in the first cohort who have ever been involuntarily job separated, the distance to the median men's sequence in the first cohort is 4.57. Therefore, in the eldest cohort the life cycle of ever-separated men is more distant from the median men work-history than that of the remainder of the sample in that cohort. Apart from the

dissimilarity index, a means contrast is reported in the last column of the table, in order to statistically check the differences by those sub-samples (EINV0 and EINV1).

In general, Table 9 shows that involuntarily job separated men are clearly more different from their respective median life-cycle sequence, except for the youngest cohort. A similar result is obtained for women: ever-separated women are comparatively more different when compared to their median sequence than their non-separated counterparts.

Therefore, involuntary job separations give way to employment status sequences which clearly differ from the median employment status sequence in each cohort. In addition, the distance to this median sequence is comparatively larger for individuals who have ever involuntarily job separated. The only exception is that of men in the fifth cohort. Our interpretation of those results is as follows. Remember that although this fifth cohort was born in the fifties, the starting average year of their first spell was 1972 (see Table 1). Given that their life course started at the beginning of the oil crises, they were presumably affected by the employment flexibility policies characteristic of the eighties. However, women in the fifth cohort do not exhibit such a modification in estimated distances when compared to the behaviour of the previous cohorts. Thus, from a historical perspective, involuntary job separations have remarkably modified men's life cycles: in particular, they have given way to sequences which are significantly different from the 'typical' life cycle of their respective cohort (i.e., the median sequence). However, for the youngest men (i.e., those who were born between 1950 and 1959) the fact of suffering an involuntary job separation at some point in time appears as more connected (or more similar) to the 'typical' life cycle.

Finally, in Table 10 we have done the same exercise as in Table 9, but, this time, considering only individuals who have 'completed' life cycles. Those individuals exit from the labour force for some definite reason (i.e., sequences ending in code 'A'). In general, the results of Table 10 for men in the second and third cohorts are quite similar to those in Table 9, with the only exception of the fifth and first cohorts. Regarding this first cohort, its few non-completed life cycles seem to be different enough so as to alter the previously-obtained measure of distance. Given that their starting average year of the first spell is 1929, individuals with completed life cycles belonging to the first cohort must presumably have been strongly affected by the Great Depression. This external influence is reflected in their respective distance measure: involuntary job separations appear more frequently. The fact that individuals ever involuntarily separated are more similar to the typical biography is also appreciated in Table 10 for the fifth cohort, which has been affected by the oil shocks (See note 33).

### **3.5. OLS regressions on sequence dissimilarities**

In this section, we provide a more comprehensive treatment of differences by estimating some ordinary least squares regressions on sequence dissimilarities. We seek to complete the results presented in the previous section, considering other factors potentially behind those distances (See note 34). The model consists of variables to control both for initial conditions and for other aspects which remain constant along time. We consider as explanatory variables the following ones: race, time elapsed since the first spell, dummies indicating the cohorts, a dummy with value one if the individual has suffered an involuntary job separation, gender, educational level —(i) higher and first degree, (ii) teaching, nursing and other, (iii) General Certificate of Education Advanced level (GCE A) or equivalent, (iv) General Certificate of Secondary Education Ordinary level (GCSE O) or equivalent and (v) vocational training—, and a dummy indicating whether or not the individual has reached the mandatory retirement age (60 for women and 65 for men).

**Table 9:** Average optimal dissimilarity index to median sequences (by gender and cohort). Total Sample

Substitution cost matrix:	Average dissimilarity index		T-test	Average dissimilarity index		T-test
Default	Men			Women		
	EINV0	EINV1		EINV0	EINV1	
Cohort 1	3.19	4.57	-4.56***	3.37	4.93	-5.76***
Cohort 2	3.42	6.15	-9.92***	5.99	6.46	-2.15**
Cohort 3	4.41	5.76	-6.74***	4.50	5.51	-3.99***
Cohort 4	2.94	6.23	-12.04***	4.25	5.34	-4.85***
Cohort 5	5.91	5.37	3.11***	4.78	6.74	-8.54***
Absolute value						
Cohort 1	2.88	4.21	-4.61***	3.08	4.62	-5.83***
Cohort 2	3.38	6.11	-10.07***	5.56	5.91	-1.76*
Cohort 3	4.09	5.24	-5.86***	4.37	5.37	-3.96***
Cohort 4	2.87	6.17	-12.01***	4.10	5.13	-4.71***
Cohort 5	5.67	4.59	6.59***	4.31	6.17	-8.41***
Data-based						
Cohort 1	2.91	4.25	-4.53***	2.95	4.39	-5.36***
Cohort 2	3.35	6.13	-10.19***	5.10	5.02	0.55
Cohort 3	3.77	5.09	-6.87***	3.83	4.88	-4.18***
Cohort 4	2.85	6.15	-11.97***	3.56	4.63	-4.93***
Cohort 5	5.62	4.43	7.58***	1.74	3.13	-6.98***

Notes:

The figures in this table indicate each cohort's optimal distance by sex to the median sequence inside each cohort (for individuals either ever involuntarily separated or not).

EINV1: Individuals ever involuntarily separated; EINV0: Individuals never involuntarily separated

\*=significant at 0.10; \*\* Significant at 0.05; \*\*\* significant at 0.01

The estimation of dissimilarity equations is showed in Table 11. The three different measures of substitution costs explained in Section 3.3 above are used to construct the mean dissimilarity that each individual has with respect to the median sequence for each cohort (those numerical values were obtained to calculate the figures showed in Table 9). Then, those three measures of individual distances are taken as dependent variables in the estimations. As can be observed, estimated distances to the median sequence in each cohort differ according to the type of distance measure considered for the analysis. However, the dummy indicating if the individual has ever been involuntarily separated is always a significant contributor to higher distances. *Ceteris paribus*, therefore, involuntary job separations provoke a larger distance to the respective median life cycle. In other words, involuntary job separations present a long-term effect on the employment history. In addition, given that this analysis considers individuals inside the same cohort, we can assert that involuntary job separations give way to histories significantly different from the median life cycle of other workers affected by the same historical context (See note 35). Considering the interaction of involuntary job separations with birth cohort, we see that involuntary job separations enlarge the distance of the fourth cohort to the respective median sequence (when compared to our reference cohort, i.e., the third one), while for the fifth cohort (the youngest one) the interaction coefficient is negative (-0.74) and almost equal to the coefficient of involuntary job separations (0.78). Therefore, the differences in labour market sequences created by involuntary job separations are closely related to the birth cohort: they are the same for the first, second and third cohort; they increase for the fourth cohort; and they are almost non-existent for the fifth cohort.

**Table 10:** Average optimal dissimilarity index to the median sequences (by sex and cohort). Only individuals with completed life cycles

Substitution cost matrix:	Average dissimilarity index		T-test	Average dissimilarity index		T-test
Default	Men			Women		
	EINV0	EINV1		EINV0	EINV1	
Cohort 1	6.96	4.18	8.89***	3.64	5.05	-5.28***
Cohort 2	3.12	5.84	-9.95***	5.25	5.56	-1.49
Cohort 3	3.81	5.90	-5.32***	4.13	4.91	-2.77**
Cohort 4	4.87	5.07	-0.38	6.86	6.08	1.00
Cohort 5	8.61	7.25	1.48	8.17	4.80	3.38***
Absolute value						
Cohort 1	6.79	4.18	8.40***	3.64	4.64	-3.71***
Cohort 2	3.11	5.84	-10.00***	4.81	5.05	-1.29
Cohort 3	3.51	5.61	-5.58***	4.02	4.79	-2.78***
Cohort 4	4.49	4.61	-0.21	6.31	5.69	0.89
Cohort 5	7.78	6.24	2.18	7.57	4.64	3.45**
Data-based						
Cohort 1	6.12	4.18	6.25***	3.61	4.04	-1.68*
Cohort 2	3.11	5.83	-9.99***	4.61	4.50	0.67
Cohort 3	3.49	5.71	-5.79***	3.59	4.19	-2.26**
Cohort 4	4.52	4.62	-0.19	6.23	5.24	1.36
Cohort 5	8.07	6.19	2.54*	7.98	4.43	3.55**

## Notes:

The figures in this table indicate each cohort's optimal distance by sex to the median sequence inside each cohort (for individuals either ever involuntarily separated or not).

EINV1: Individuals ever involuntarily separated; EINV0: Individuals never involuntarily separated

\*=significant at 0.10; \*\* Significant at 0.05; \*\*\* significant at 0.01

Regarding gender, this dummy variable (itself or its interaction with involuntary job separations) does not significantly affect dissimilarity indices. Therefore, gender is not a key variable (in the present) to detect significant changes from the median employment status sequence in each cohort. We consider this to be a result arising from the increasing similarity of male and female histories detected in previous sections (See note 36).

Finally, we comment the results for cohort variables. Apparently, the results (specially, the sign of coefficients) are not reflecting the cohort patterns described in previous sections. However, interpretation of those regression results must be different from the description given in previous sections, given that distances used as dependent variables in Table 11 have been calculated inside each cohort. That is, cohort dummies in this table capture the impact—compared to our reference cohort, i.e, the third one—that the fact of belonging to each cohort presents on the distance to the median sequence *inside* each cohort. If we stick to the data-based substitution cost matrix, from the interactions between cohort and the separation dummy we can assert that individuals in the fourth cohort suffering involuntary job separations are significantly more different from the median sequence. In addition this estimated coefficient (0.44) compensates for the negative one in the fourth cohort dummy (–0.41). For the fifth cohort, suffering involuntary job separations reduces the estimated distance to the median sequence. Therefore, a jointly consideration of the estimated coefficients for the cohort dummies and their interactions with involuntary job separations leads us to obtain a similar conclusion to the previous sections (i.e., that mobility is a comparatively more ‘natural’ event for the youngest cohorts than for the remainder of the sample).



**Table 11:** OLS Estimates for Dissimilarity Indices

Variable	SUBSTITUTION COST MATRIX					
	Absolute Value N=6159		Default N=6159		Data-Based N=6159	
	Coeff.	T-ratio	Coeff.	T-ratio	Coeff.	T-ratio
Race (Whites=1)	.02	0.08	.02	0.11	.07	0.35
Years since the first spell	-.01	-1.44	-.01	-1.33	-.01	-2.59***
Cohorts:						
Cohort1	-.82	-3.72***	-.71	-3.10***	-.55	-2.56
Cohort2	.66	3.36***	.84	4.10***	.59	3.13***
Cohort4	-.40	-2.29**	-.46	-2.52**	-.41	-2.39**
Cohort5	.33	1.71*	.53	2.65***	.38	2.04**
Involuntary job separation	.82	4.77***	.81	4.59***	.78	4.70***
Gender (Males=1)	.01	0.07	.06	0.35	-.01	-0.08
Education:						
Higher and first degree	.17	1.26	.24	1.68*	.16	1.20
Teaching, nursing and other	.06	0.68	.05	0.53	.04	0.46**
GCE A level or equivalent	-.01	-0.05	-.02	-0.15	-.08	-0.56
GCSE O level or equivalent	.02	0.25	.01	0.13	.04	0.44
Vocational training	-.10	-0.82	-.11	-0.88	-.07	-0.58
Mandatory retirement age	.05	0.28	.02	0.14	.17	0.98
Interactions:						
Invol. Separated*Cohort1	.13	0.51	.19	0.69	.09	0.34
Invol. Separated*Cohort2	.30	1.34	.34	1.47	.17	0.77
Invol. Separated*Cohort4	.44	2.11**	.43	1.99**	.44	2.14**
Invol. Separated*Cohort5	-.56	-2.70***	-.38	-1.74*	-.74	-3.67***
Interactions:						
Invol. Separated*Gender	-.01	-0.07	.05	0.35	.09	0.67
Interactions:						
Gender*Cohort1	-.12	-0.50	-.23	-0.89	-.03	-0.13
Gender*Cohort2	-.73	-3.31***	-1.07	-4.71***	-.46	-2.16**
Gender*Cohort4	-.23	-1.05	-.34	-1.52	-.06	-0.30
Gender*Cohort5	.16	0.75	.02	0.10	.29	1.41
Constant	4.59	14.22***	4.79	14.39***	4.37	13.96***
R <sup>2</sup>	0.06		0.07		0.06	
F-value	18.47***		20.78***		16.85***	

Notes: Reference individual is woman, not white, belongs to Cohort 3, has never been involuntarily-job separated, has CSE, Scot Grade or no qualification, and has not yet reached the mandatory retirement age (60 for women and 65 for men).

\* = Significant at 0.10 level

\*\* = Significant at 0.05 level

\*\*\* = Significant at 0.01 level or better

## 4. Conclusions

We have used work-life history data from the British Household Panel covering from individual's first job to 1993 in order to perform an empirical analysis of the

employment status mobility from a life-cycle perspective. Involuntary job separations, gender and birth cohort are the three variables on which our research focus has centered. We make two principal contributions. First, we have used an alternative methodological approach (optimal matching analysis) to demonstrate the existence of a systematic impact of involuntary job separations on life course mobility. This procedure allowed us to better understand the factors that influence the mobility across various stages of the life course for five different cohorts.

Our second contribution has been to provide new information about historical employment mobility changes in Britain, placing a special focus on the incidence and consequences of involuntary job separations. Several main findings are worth underlying from this empirical sequence analysis:

- Older cohorts have fewer spells and less changes of employment status between successive spells. On the contrary, younger cohorts have more spells, they are more mobile —although their greater mobility does not include job-to-job movements— and present more homogeneous histories. Therefore, mobility in employment status has increased along the twentieth century.

- The historical change in female participation in the labour market would help to explain this increase in mobility, given that women sequences have become more similar to those of men.

- Birth cohorts in the second half of the century have been comparatively more affected by involuntary job separations, when compared to the eldest cohorts.

- Applying OMA techniques we find that there is a significant difference in employment status sequences between those individuals who have ever been involuntarily job separated and those who have not. In particular, involuntary job separations give way to employment status sequences which substantially differ from the median sequence in each cohort, however this effect is not constant across cohorts (and it is very small for the fifth cohort).

- Differences in sequences obtained applying OMA show that male employment sequences are more affected by involuntary job separations in the youngest cohort. That is, experiencing involuntary job separations becomes more ‘natural’ for men along the seventies. However, the gender differences are not confirmed in the regression analysis, corroborating that (*ceteris paribus*) women sequences are equal to those of men.

On the whole, these findings provide a changing picture of employment status mobility in Britain along the twentieth century. Mobility has increased in the second half of the

century and is now more affected by involuntary job separations. Evidence on a historical change in labour market participation of women is shown through increasing similarity between men and women employment status sequences. Therefore, the approach followed in this article has provided relevant (and previously unknown) information about the historical changes of employment status mobility in Britain, in special underlying the increasing importance of involuntary job separations.

## **5. Acknowledgements**

The authors would like to thank the comments and suggestions from two anonymous referees, which proved very useful to improve the final version of the article. This research is partially based on work carried out during a visit of Miguel A. Malo to the European Centre for Analysis in the Social Sciences (ECASS) at the University of Essex. The data used in this paper were made available through the ESRC Data Archive. The data were originally collected by the ESRC Research Centre on Micro-Social Change at the University of Essex. Neither the original collectors of the data nor the Archive bear any responsibility for the analyses or interpretations presented here. Miguel A. Malo also acknowledges the financial support received from the European Commission. The usual disclaimer applies.

## Notes

1. See Mincer and Polachek (1974 and 1978), Corcoran and Duncan (1979), Goldin (1989), or Hill and O’Neill (1992).
2. See, for instance, Jacobson et al. (1993) for the US and Elias (1997) for the UK.
3. See, for instance, Hildreth *et. al* (1998) or Booth *et. al* (1999), who also use data from the British Household Panel Survey.
4. See, in this respect, Abbot (1985), and Abbot and Forrest (1986).
5. Individuals under 34 years-old at 1-December-93 were also initially included in the analysis. Nonetheless, due to the short length of their observed work-histories, no significant results were attained regarding this sixth cohort. Therefore, they were excluded from the empirical analysis.
6. See Halpin (1997a) and Taylor (1996) for more complete details about the BHPS.
7. Following Halpin (1997a) suggestions, we have used the weights supplied in the main database. As the job history information is included in our data files, the weight variable used has been CXEWGHT which selects the survey respondents and allows to adjust the sample selection and attrition effects. We have preferred the selection of survey *respondents*, because the proxies could not give accurate information on employment and job histories. On the weights variables in the BHPS and how to use them, see Taylor (1997).
8. See Elias (1991) and Elias (1997a).
9. As we will see, our findings on employment status mobility patterns can also be interpreted in the light of the historical process of the increasing participation of females in the labour market, in addition to those exogenous macroeconomic phenomena. This fact stresses, again, the importance of analyzing the evolution along time of the different birth cohorts.
10. The usual mobility measures in the literature are net mobility indicators. That is, they compare employment status in two different states (Elias, 1997b). On the other hand, the two measures we propose for mobility allow to attain a vision of the life cycle in the long-term.
11. We have divided the information provided by the survey to the question about reasons to leave the job in two groups. Firstly, involuntary reasons: made redundant; dismissed or sacked; temporary job ended; or stopped by health reasons.

Secondly, the remainder of reasons (presumably voluntary): promoted/left for a better job; took retirement; left to have a baby; and children/home care.

12. As it is well-known, very complex element schemes may retain important substantive information, but they also increase the computational intensity of the analyses, and may make it difficult to identify similar sequential patterns; conversely, overly simple schemes may disguise meaningful variation in sequential pattern.
13. Its rank is between 0 and 38 and it must be always lower than the number of spells of the individual.
14. Moreover, the gross mobility measure seems to decrease for the youngest cohorts. However, the shortest time period considered for those cohorts might be explaining this relationship.
15. This is a straightforward tool for studying labour market attachment: since we dispose of a set of discrete events, we can write down all of their logical combinations and then count the frequency of their occurrence in each sample (Berger, Steinmuller and Ziegler, 1993, and Hogan, 1978).
16. Instead, previous research has detected an under-reporting of unemployment spells, especially the shortest ones —see, for instance, Dex and McCulloch, 1997, and Elias, 1997a.
17. Given that this highly unlikely event would constitute the main argument in order to be able to assimilate this second group to the pattern of the eldest cohorts described above, this exploratory analysis allows us to state that this second group might be showing a new pattern of employment status mobility.
18. At least in their first spell, since the order of sequences to obtain medians is lexicographic.
19. The most frequent sequences up to four spells represent almost 50 per cent for the fifth cohort, while for the eldest cohorts they only represent less than 25 per cent.
20. See Sankoff and Kruskal, 1983.
21. Abbot and Forrest, 1986.
22. Abbot, 1995; Abbot and Hrycak, 1990; Stovel, Savage and Bearman, 1996.
23. Chan, 1994; Halpin and Chan, 1998.

24. Baizán, Michielin and Billari, 2002; Solís and Billari, 2002. See also Billari (2001) and Billari and Piccareta (2001) for respective discussions of the sequence analysis approach to the study of life courses.
25. See Isaac *et al.* (1991) and Box *et al.* (1994) in this respect.
26. Useful reviews are found in Bondon (1973) and Sterman (1976).
27. A review of this method is found in Yamaguchi, 1991.
28. Appendix B gives a full account of the mathematical algorithm used by optimal matching.
29. Symmetry does not automatically hold, but can be forced by equating insertion and deletion costs, or by a slightly different definition: minimum cost of transforming  $a$  into  $b$  or  $b$  into  $a$ . Whether the distance measure will also be transitive, and thus constitutes a metric distance, depends on the definition of the set of elementary operations,  $\mathbf{Z}$ . Transitivity is automatically guaranteed for the simple case when there are only insertions, deletions and substitutions.
30. In many applications, the cost of insertions and deletions are fixed at a value slightly higher than the highest substitution cost (see Abbott and Hrycak, 1990).
31. Insertion and deletion costs may themselves vary. However, these costs are in some sense a function of what the sequence already looks like; because of this uncertainty, most applications make indel costs the same in all cases (see Abott and Hrycak, 1990, pp. 155).
32. The relationship is as follows: length of longest common subsequence= $0.5*(m+n-d(a,b))$ , where  $d(a,b)$  is the optimal matching distance when using the cost functions as indicated in the text;  $m$  and  $n$  denote the length of the first and second sequence, respectively.
33. Results for the fourth and fifth cohorts must be cautiously analysed because their completed life cycles must be rather different from the completed life cycles of the remainder cohorts. Due to their limited sample size, only occasionally the average dissimilarity measures are significantly different for individuals ever and never involuntarily job separated in those two youngest cohorts.
34. McVicar *et. al* (2001, pp. 325) also make use of regression techniques with the output of optimal matching analysis, and indicate other papers adopting a similar approach.
35. The distance gap between ever and and never involuntarily job separated workers can be decomposed into a portion “explained” by differences in characteristics and

a portion “unexplained” following the decomposition by Blinder (1973) and Oaxaca (1973). This allows us to test whether or not the effect of the remainder explanatory variables on the dissimilarity indices differs between those never involuntarily job separated and those ever involuntarily job separated. We have estimated separately OLS estimations for both groups, and have worked out the Oaxaca-Blinder decomposition. In addition, given that individuals might not be randomly distributed to those groups, we have tested whether or not there is selection by workers into both subgroups. A two-stage Heckman procedure was applied for this purpose. The most relevant result is that having been involuntarily job separated is by far the chief factor explaining individuals’ distance to the median sequences in each cohort. Those results are not shown, but are available from the authors upon request. Moreover, it is observed that distances to the median sequence are almost unrelated to the explanatory variables. This fact may explain the low  $R^2$  values showed in Table 11. That is, the variables introduced in the model—in order to control for initial conditions and for other aspects which remain constant along time— explain less than 10 percent of the individuals’ work-history. The remainder 90 percent is then explained by other factors, such as luck, the historical moment, and the decisions that each individual takes along his/her life-course.

36. We have estimated the regressions without interactions and the estimated coefficient for gender present a significantly negative effect—with the exception of those estimations based on the absolute value substitution cost matrix. Therefore, being male would decrease the distance to the median sequence in each cohort, although this effect is compensated for by the interaction with involuntary job separations.

## References

- Abbot, A. (1985). "Sequence Analysis: New Methods for Old Ideas", *Annual Review of Sociology*, 21: 93-113.
- Abbot, A. and J. Forrest (1986). "Optimal Matching Methods for Historical Sequences", *Journal of Interdisciplinary History*, 16: 471-494.
- Abbot, A. and Hrycak, A. (1990). "Measuring Resemblance in Sequence Data: An Optimal Matching Analysis of Musicians' Careers". *American Journal of Sociology*, 96(1): 144-185.
- Baizán, P., F. Michielin, and F.C. Billari (2002). "Political Economy and Life Course Patterns: The Heterogeneity of Occupational, Family and Household Trajectories of Young Spaniards", *Demographic Research*, vol. 6, art. 8, 192-240.
- Berger, P.A., Steinmuller, P. and Ziegler, Z. (1993) "Upward Mobility in Organisations: The Effects of Hierarchy and Opportunity Structure". *European Sociological Review*, 9:173-188.
- Billari, F.C. and R. Piccarreta (2001), "Life courses as sequences: an experiment in classification via monothetic divisive algorithms", in S. Borra, R. Rocchi, M. Vichi, M. Schader (Eds.), *Advances in Classification and Data Analysis*, Springer Verlag, Berlin and New York: 351-358.
- Bondon, R. (1973). *Mathematical Structures of Social Mobility*, San Francisco: Jossey Bass.
- Booth, A.L., M. Francesconi, and C. García-Serrano (1999). "Job Tenure and Job Mobility in Britain", *Industrial and Labor Relations Review*, vol. 53, no. 1, 43-70.
- Box, G.E.D, Jenkins, G.M. and G.C. Reisel (1994). *Time Series Analysis*, Englewood Cliffs, NJ: Prentice Hall.
- Chan, T.W. (1994). *Tracing Typical Mobility Paths*. Ms. Nuffield Coll. Oxford.
- Chan, T.W. (1995). "Optimal Matching Analysis: A Methodological Note on Studying Career Mobility", *Work and Occupations*, 22 (4), 467-490.
- Corcoran, M. and Duncan, G. (1979): "Work History, Labor Force Attachment, and Earnings Differences Between the Races and the Sexes", *Journal of Human Resources*, 14 (Winter), pp. 3-20.



- Creedy, J. and Disney, R. (1981): “Changes in Labour Market States in Great Britain”, *Scottish Journal of Political Economy*, vol. 28, no. 1, February, pp. 76-85.
- Dex, S. and McCulloch, A. (1997): “The Reliability of Retrospective Unemployment History Data”, *Working papers of the ESRC Research Centre on Micro-Social Change*, paper 97-17 Colchester, University of Essex.
- Elias, P. (1991). “Methodological, Statistical and Practical Issues Arising from the Collection and Analysis of Work History Information by Survey Techniques”, *Bulletin de Methodologie Sociologique*, no. 31, 3-31.
- Elias, P. (1997a): “Who forgot they were unemployed?”, *Working papers of the ESRC Research Centre on Micro-Social Change*, paper 97-19 Colchester, University of Essex.
- Elias, P. (1997b): “Restructuring, Reskilling and Redundancy: A Study of the Dynamics of the UK Labour Market, 1990-95”, *Working papers of the ESRC Research Centre on Micro-Social Change*, paper 97-20 Colchester, University of Essex.
- Farber, H.S. (1994), "The Analysis of Interfirm Worker Mobility", *Journal of Labor Economics*, 12, 554-594.
- Goldin, C. (1989): “Life-Cycle Labor Force Participation of Married Women: Historical Evidence and Implications”, *Journal of Labor Economics*, vol. 7, no. 1, pp. 20-47.
- Halpin, B. (1997a): “Unified BHPS Work-life Histories: Combining Multiple Sources into a User-friendly Format”. *Technical Papers of the ESRC Research Centre of Micro-Social Change*. Technical Paper 13. Colchester, University of Essex.
- Halpin, B. (1997b): “Is Class Changing? Evidence from BHPS Retrospective Work-life Histories”, *Working papers of the ESRC Research Centre on Micro-Social Change*, paper 97-13 Colchester, University of Essex.
- Halpin, B. and T. W. Chan (1998) Class Careers as Sequences: an Optimal Matching Analysis of Work-Life Histories , *European Sociological Review*, 14: 113-130.
- Hill, M.A. and O'Neill, J.E. (1992): “Intercohort Change in Women’s Labor Market Status”, *Research in Labor Economics*, vol. 13, pp. 215-286.
- Hildreth, A. K., S.P. Millard, D.T. Mortensen and M. Taylor (1998). “Wages, Work and Unemployment”, *Applied Economics*, vol. 30, 1531-1547.
- Hogan, D.P. (1978): “The Variable Order of Events in the Life Course”, *American Sociological Review*, 43: 573–586.

- Isaac, L., S.M. Carlson and M.P. Mathis (1991). *Quality of quantity in comparative historical time series analysis*, Presented at Soc. Sci. Hist. Assoc., New Orleans, 2, November.
- Jacobson, L.S., R.J. Lalonde and D.G. Sullivan (1993): “Earning losses of displaced workers”, *American Economic Review*, 83, 4, pp. 685-709.
- McVivar, D., and M. Anyadike-Danes (2002): “Predicting successful and unsuccessful transitions from school to work by using sequence methods”, *Journal of the Royal Statistical Association, Series A*, 165, Part 2, pp. 317-334.
- Mincer, J. and S. Polachek (1974): “Family Investments in Human Capital”, *Journal of Political Economy*, 82 (March/April), pp. S76-S108.
- Mincer, J. and S. Polachek (1978): “Women’s Earnings Reexamined”, *Journal of Human Resources*, 13 (winter), pp. 118-133.
- Sankoff, D. and J.B. Kruskal (1983). *Time Warps, String Edits and Macromolecules*, Reading, MA: Addison-Wesley.
- Solís, F. and F.C. Billari (2002). “Work Lives amid Social Change and Continuity: Occupational Trajectories in Monterrey, México”, Max Planck Institute for Demographic Research, working paper WP 2002-009.
- Stewman, S. (1976). “Markov Models of Occupational Mobility”, *J. Math. Soc.*, 4: 201-45, 247-78.
- Stovel, K., M. Savage and P. Bearman (1996). “Adscriptio into Achievement: Models of Career Systems at Lloyds Bank, 1890-1970”, *American Journal of Sociology*, 102: 358-399.
- Taylor, M.F. (1997): “British Household Panel Survey User Manual”, User documentation, ESRC Research Centre on Micro-Social Change.
- Wu, Lawrence L. (2000). “Some Comments on Sequence Analysis and Optiaml Methods in Sociology”, *Sociological Methods and Research*, 29 (4): 41-64.
- Yamaguchi, K. (1991). *Event History Analysis*, Newbury Park, CA: Sage.

## Appendix A.

**Table A.1:** Files in the BHPS database used for the empirical analysis

Filename	Wave	Start of field work	Description
AINDRESP	1	Sept 1991	The main individual respondent file, containing <i>inter alia</i> detailed information on current status at the date of interview
AJOBHIST	1	Sept 1991	Information on all employment status spells between 1/9/90 and the date of interview
BINDRESP	2	Sept 1992	Wave 2 equivalent of AINDRESP
BJOBHIST	2	Sept 1992	Inter-wave history: details of all employment status spells between 1/9/91 and the date of interview.
BLIFEMST	2	Sept 1992	Information on all employment status spells since first leaving full-time education until the date of interview
CLIFEJOB	3	Sept 1993	Information on all jobs held since first leaving full-time education until the beginning of data collection

**Table A.2:** Main descriptive statistics

Variable	Mean	Std. Deviation	Minimum	Maximum
Race (Whites=1)	.98	.15	0	1
Years since first spell	38.79	16.46	.33	87.25
Cohort1	.15	.35	0	1
Cohort2	.19	.40	0	1
Cohort4	.24	.43	0	1
Cohort5	.24	.43	0	1
Involuntary job separation	.33	.47	0	1
Gender (Males=1)	.44	.50	0	1
London	.11	.32	0	1
Gales	0.05	.22	0	1
Scotland	0.07	.27	0	1
Higher and first degree Education	0.07	.26	0	1
Teaching, nursing and other Education	.19	.39	0	1
GCE A level or equivalent Education	0.06	.25	0	1
GCSE 0 level or equivalent Education	.16	.36	0	1
Vocational training Education	0.07	.27	0	1
Mandatory retirement age	.36	.48	0	1

## Appendix B. Basic Optimal Matching Algorithm

If the elementary operations consist only of insertions, deletions and substitutions, the distance measure can be calculated with a simple dynamic programming method. To explain this method, we mainly follow Sankoff and Kruskal (1983, pp. 266).<sup>1</sup>

Let  $\mathbf{Y}$  denote the finite state space and  $\phi$  an “empty state” which is not contained in  $\mathbf{Y}$ . The cost functions are:

$$\begin{aligned} \psi(x,y) &= \text{substitution cost, } x,y \in \mathbf{Y} \\ \psi(x, \phi) &= \text{deletion cost, } x \in \mathbf{Y} \\ \psi(\phi, y) &= \text{insertion cost, } y \in \mathbf{Y} \end{aligned}$$

The expression:

$$\begin{bmatrix} x_1 & \dots & x_p \\ y_1 & \dots & y_p \end{bmatrix} \tag{1}$$

is called an *alignment* if  $x_i, y_i \in \mathbf{Y} \cup \{\phi\}$  and there is no column with  $x_i = y_i = \phi$ . The *length* of an alignment (i.e., the distance between the two sequences) is defined by

$$\sum_i w(x_i, y_i) .$$

Now, let  $a=(a_1, \dots, a_m)$  and  $b=(b_1, \dots, b_n)$  be two sequences with states in  $\mathbf{Y}$ . We say that (1) is an alignment between the sequences  $a$  and  $b$  if by inserting empty states into  $a$  it can be made equal to  $x=(x_1, \dots, x_p)$ , and in the same way  $b$  can be made equal to  $y=(y_1, \dots, y_p)$ . Given this definition, we can finally define the (standard) distance between  $a$  and  $b$ , denoted  $d(a,b)$ , as the minimum possible length of any alignment of these two sequences. Additional notation to describe the algorithm is:  $a^i=(a_1, \dots, a_i)$ ,  $b^j=(b_1, \dots, b_j)$ , and  $d_{ij}=d(a^i, b^j)$ .

---

<sup>1</sup> Computations for this algorithm have been implemented through the use of the computer program called TDA (*Transition Data Analysis*), which is publicly available at the following web site: <http://steinhaus.stat.ruhr-uni-bochum.de/tda.html>

Using this notation, we need to calculate  $d(a,b)=d_{mn}$ . This can be done by calculating the elements of the  $(m+1,n+1)$  matrix  $D= (d_{ij})$  ( $i=0,\dots,m$ ;  $j=0,\dots,n$ ) recursively in the following way. The first step is initializing the first row and first column of this matrix:

$$\begin{aligned} d_{0,0} &= 0 \\ d_{0,j} &= d_{0,j-1} + w(\phi, b_j) \quad j=1,\dots,n \\ d_{i,0} &= d_{i-1,0} + w(a_i, \phi) \quad i=1,\dots,m \end{aligned}$$

All other elements of  $D$  can then be calculated by using three predecessors. The recurrence relation is:

$$d_{ij} = \min \{ d_{i-1,j} + w(a_i, \phi), d_{i-1,j-1} + w(a_i, b_j), d_{i,j-1} + w(\phi, b_j) \}$$

Having calculated all elements of  $D$ , one finally finds the required distance in the element  $d_{mn}$ .

