DURATION MODELS AND DIFFERENTIAL EVOLUTION IN THE ANALYSIS OF LARGE DATA SETS

Michele Lalla[*], Sandra Paterlini

Department of Political Economy, Via Jacopo Berengario 51, 41100 Modena, Italy

Abstract

Two measures of unemployment duration were obtained for each individual from data collected by Istat through the Quarterly Labour Force Survey. One measure was the length of on-going spells of unemployment, declared by the interviewed individual in a first wave and suitably adjusted considering information in a second wave (Spell Answers). The other was the time elapsed between the dates of the end of the previous job and the beginning of the current job in the second wave (Spell Reconstructions). They were modelled by traditional duration models via a mixture distribution of the answered and reconstructed spells, instead of using their minimum. The likelihood of the mixture is highly complex with a large number of covariates. Hence, classical optimisation methods do not provide robust parameter estimations. A stochastic search algorithm, called Differential Evolution, is proposed to tackle the numerical optimisation problem. In fact, even if DE is still rather unknown, it proves to be consistently and clearly superior to other search heuristic and classical optimisation approaches in many applications. Moreover, it is very simple to implement and requires little or no parameter tuning. A comparison between the ordinary duration models based on a sole unemployment spell for individuals (spell answers or spell reconstructions) is also reported.

JEL classification: C41, J64

Keywords: Unemployment duration, unobserved heterogeneity, heaping effect, Differential evolution, proportional hazard model

1. Introduction

The average length of time that workers spend seeking a job is an important indication of economic welfare, but, despite this role, its measurement still appears elusive, as is the question: How long do spells of unemployment last? One reason is almost an ontological issue, as it lies in the addition of the fourth

[*] Corresponding author.

E-mail address: lalla.michele@unimore.it

dimension to the sampling strategy: survey populations constantly change over time in the composition and the characteristics of their members, giving rise to many problems and different objectives for analysis (Duncan and Kalton, 1987). There are a variety of sample designs for collecting data at several points of time. Each of them varies in the amount and kind of overlaps between periods, *i.e.*, the percentages of the same statistical units to retain in the sample for each wave, or round, of data collection and the number of waves. The Italian Quarterly Labour Force Survey (QLFS) is conducted by Istat and uses a rotation scheme 2-2-2 for the sampling units, which are families. Each rotation group remains in the survey for two consecutive quarters, drops out for two subsequent quarters, and then returns for two more consecutive quarters, as described in Table 1. Each survey consists of a point-in-time data collection, as is the Current Population Survey (CPS), and it cannot measure important aspects of phenomena enduring over time (Kiefer et al., 1985). However, the rotation scheme acts firstly to maintain an up-to-date sample of a changing population, secondly to reduce panel conditioning and panel loss with respect to the nonrotating panel survey, and thirdly to detect the transitions between the labour market states of the individuals interviewed in two different waves, t_1 and t_2. For example, in Table 1, group C was interviewed in 1998:1 and 1998:2, or group E in 1997:4 and 1998:3 (Torelli and Trivellato, 1988).

Table 1
Rotating panel design of the Italian Quarterly Labour Force Survey

Year	Q	Month	Rotation groups										
1997	4	October	A	B			E	F					
1998	1	January		B	C			F	G				
1998	2	April			C	D			G	H			
1998	3	July				D	E			H	I		
1998	4	October					E	F			I	L	
1999	1	January						F	G			L	M

The ordinary transitions in the labour market considered referred to three states (professional conditions): Employed (E), Unemployed (U), Out-of-the-Labour-Force (OLF). The duration of unemployment was the time elapsed between two dates for an individual *i*, subject to some specific conditions: (a) at least fifteen years old, (b) without work during the reference week, (c) available to start work or to accept a job offer before the end of the next two weeks, (d) «actively» searching for an occupation. He/she begins the action at a certain point in time, $t_{B;i}$, and continues until the end, at another point in time, $t_{E;i}$, because he/she finds a job or he/she reduces the «intensity» of the search, thus leaving the labour force. The unemployment spell for the *i*-th individual would be the random variable $T_i = t_{E;i} - t_{B;i}$. However, it not easy or convenient to detect the intensity of the search over time and condition (d) is collected at the date of survey, ascertaining that the interviewee carried out at least one job search attempt in the previous four weeks, with respect to the reference week. The situation is illustrated in Fig. 1, where calendar time is the abscissa, the lower bound of a horizontal segment indicates the time when the spell started, the upper bound shows when it concluded. The length of each segment is an observation of a random variable, termed *completed spell* of unemployment. However, almost all spells will be in progress when each survey is conducted, i.e., the durations collected by the Italian QLFS are *incomplete spells* of unemployment, as in an ordinary CPS. In Fig. 1, case c_1 represents the standard situation, where all data about employment duration are available. The other cases illustrate only some situations with missing data or nonsampling variable errors, such as processing and field errors in coding and tabulating components. For example, c_2 denotes a direct transition into the employment state, which is generally not collectable, c_3 refers to the unavailable date of the beginning of an unemployment spell indicated by a question mark (but limited by the time interval elapsing between two consecutive waves at t_1 and t_2), c_4 concerns a right-censored spell, and so on.

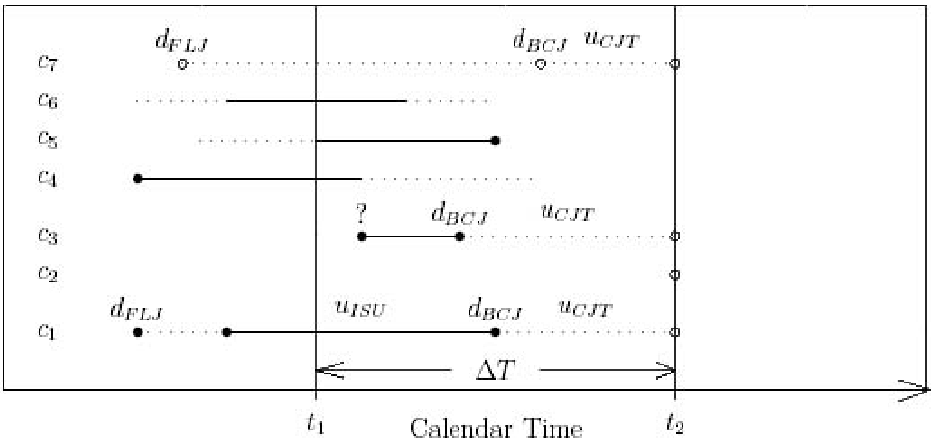

Fig. 1 – Different types of spells collectable between two generic waves conducted at t_1 and t_2, respectively.

The analysis of duration data through the hazard function has become a standard technique in the last two decades and it is currently used in many scientific and social fields, although the terminology changes from one discipline to another, thus known as reliability theory, survival analysis, and duration data

(Kalbfleisch and Prentice, 1980; Cox and Oakes, 1984; Lancaster, 1990). Analyses of the labour market and turnover through the life table method date back to Silcock (1954), but this method received more attention after the applications of Nickell (1979a,b) and Lancaster (1979) and has been extensively developed in recent decades. However, model building and parameter estimation can be difficult tasks in the analysis of labour market transitions, when the distribution function of duration is not easy to handle, all spell lengths are censored, and the number of explanatory variables (covariates) is extremely high. In the absence of these problems, unemployment spells can be handled through highly standardized methods encompassing almost routine procedures (Kleinbaum, 1996; Lee, 1993), but the data modelled below present two specific complications, namely a complex distribution function and a large set of covariates.

The standard definition of unemployment duration is given by the minimum between two spells: (i) job search duration, and (ii) length of the period since the last job held (EUROSTAT, 1996).[1] Therefore, the completed spell of unemployment is an unobserved variable for which two measures could be available. One is the length of time spent in seeking an occupation, declared by interviewed individual, surveyed as an incomplete spell, referred to as *spell answer* (SA), and denoted by the random variable T_A. Another is the length of time since last employment, which is reconstructed starting from the dates of the end of the last job and the beginning of the current job, referred to as *spell reconstruction* (SR), and denoted by the random variable T_R.

The aim of the paper is to analyse unemployment spells using traditional duration models, attempting to use both of these two unemployment spells simultaneously, combined in a mixture of distributions. Generally, researchers should choose only one measure, which could be one of them, or their minimum. To account for unobserved heterogeneity and heaping effect, the likelihood of the mixture model becomes highly complex with a large number of explanatory variables. Hence, a robust estimation of the parameters in duration models is difficult to achieve using classical optimisation methods. We propose to use a stochastic search algorithm, called Differential Evolution (DE), to tackle the numerical optimisation problem. In fact, even if DE is still rather unknown, it proves to be consistently and clearly superior to other search heuristic and classical optimisation approaches in many applications (Storn and Price, 1995). Moreover, DE is very simple to implement and requires little or no parameter

[1] A new definition of unemployment (EC Commission Regulation No. 1897/2000) may cause a break between 2000 and 2001, but it is fully compatible with ILO standards and increases comparability in the EU, not yet fully applied in Italy (EUROSTAT, 2002).

tuning. Comparison with duration models on the single measures, SA and SR, is also reported in the following.

For the sake of brevity, the transitions considered in the following refer to the movement of individuals from the unemployment state to the employment state. The paper proceeds as follows. Section 2 illustrates the available data and the procedure used to determine the two measures of unemployment duration. Section 3 outlines Cox's Proportional Hazards Model (CPHM) with a Weibull specification of the baseline hazard, the inclusion of the unobserved heterogeneity, the heaping effect, and the likelihood functions of the models. Section 4 contains a brief description of the DE stochastic search algorithm. Section 5 describes the covariates introduced in the model and the estimation of parameters. Section 6 concludes the paper with some comments and remarks.

2. Sampling data

The unemployment durations modelled below were obtained from six waves of the Italian QLFS (1997:4–1999:1), as illustrated in Table 1, *i.e.*, six databases were available, starting from the fourth quarter of 1997 to the first quarter of 1999. They cover a period of 18 months and each file contains more than 200,000 individuals. The choice of the investigation period was arbitrary, but such that it was long enough to determine the complete unemployment spells and short enough to avoid nonstationary behaviour and the possible presence of economic shocks.

For each job seeker, the questionnaire used by the QLFS provides two elements from which it is possible to evaluate unemployment spell duration. One concerns the answer to the Question: How many months have you been looking for a new job? (Q46). The spell answer ranges from 1 to 99 months and it is the length of the on-going (incomplete) spell of unemployment, as the respondent is currently unemployed, u_{ISU}. Therefore, the completed spell of unemployment should be determined considering the transition that took place between two surveys (waves) and the length of current job tenure, u_{CJT}, obtained in the second wave by the answer to the question: When did you begin your current job? (Q24, which requires specification of the month and year, d_{BCJ}). Hence, the completed spells for transitions from unemployment to employment is given by: $T_A = u_{ISU} + 1_{CJT}(\Delta t - u_{CJT}) + (1 - 1_{CJT})\Delta t / 2$, where T_A denotes the so-called spell answer (SA), Δt is the time interval between the two linked surveys detecting transition occurrence (see Fig. 1), 1_{CJT} is an indicator variable equal to 1 when u_{CJT} is not missing and less than Δt. Otherwise, 1_{CJT} is equal to 0. When u_{CJT} is greater than Δt, presumably there is an error in the data.

The other evaluation of unemployment duration considers the time elapsed between the dates of the end of the previous job and the beginning of the current job. The former is generally obtained in the first wave by the answer to the question: When did your last job end? (Q38, which requires specification of the month and year, d_{FLJ}). The latter is usually given by d_{BCJ}, defined previously. Therefore, for individuals with changes in the unemployment state between two waves, it is possible to determine the time elapsed between two jobs which could be an approximate measure of an unemployment spell: $T_R = g(d_{BCJ}, d_{FLJ})$, where T_R denotes the so-called spell reconstruction (SR) and $g(\cdot)$ is a function computing the length (in months) of a period delimited by two dates.

A possible impreciseness of T_R lies in the practical process of job searching. In fact, job losers/leavers could remain out-of-the-labour-force because they are not looking for work after losing/leaving job. Therefore, in abstract terms, T_R should be greater than or equal to T_A, but the two types of unemployment spells are not always available (missing data) and T_R is not always greater than or equal to T_A. However, T_R may also assume values that are too high —$t_{R;i}$ was greater than 120 months for 457 out of 15123 individuals with T_R being available (3.5%), where the linked and selected units amounted to 20617—, while T_A ranged between ½ to 108 months involving 13309 individuals. For the 457 individuals with $t_{R;i} > 120$ months, $t_{A;i}$ was never missing and only 26.9% of them answered «99 months» to Q46. However, the latter were eliminated from the sample and the number of cases became 20160. For this reduced data set, the distribution of differences between the two durations, $d_i = t_{A;i} - t_{R;i}$, is reported in Table 2: it showed a positive mean ($\bar{d} = 4.1$ months), skewed to the right ($g_1 = 0.368$), and also proved to have a pronounced leptokurtosis ($g_2 = 3.792$). The latter apparently implies the contrary of what was expected, *i.e.*, $t_{A;i}$ is more frequently greater than $t_{R;i}$. However, with the inclusion of the 457 individuals eliminated, the conclusions overturned: $\bar{d} = -4.7$ months, $g_1 = -3.368$, and $g_2 = 22.220$, as expected.

There are nine possible transitions between the three professional conditions (E, U, OLF), but some of them generate spell durations that are incomplete and/or over-represent short lengths, such as transitions from employment to employment or from OLF to OLF, while others are not of direct interest here, such as those from employment to OLF for retirement. The selected transitions are reported in Table 3, together with distinctions as to the characteristics of u_{CJT} used to determine $t_{A;i}$.

Table 2
Absolute frequency and percentage distribution of the differences between the answered and reconstructed spells ($d_i = t_{A;i} - t_{R;i}$) with the lower (LB) and upper bounds (UB) of the class intervals (the number of missing values is 12,712)

LB **UB**	$-\infty$ –10.5	–10.5 –5.5	–5.5 –2.5	–2.5 –1.5	–1.5 –0.5	–0.5 0.5	0.5 1.5	1.5 2.5	2.5 5.5	5.5 10.5	10.5 $+\infty$	**Total**
n_i	1019	263	288	164	672	1638	549	230	439	440	1746	**7,448**
%	13.7	3.5	3.9	2.2	9.0	22.0	7.4	3.1	5.9	5.9	23.4	**100.0**

Table 3
Number of individuals and column percentages for the transitions detected between two waves

Transitions	Conditions	Employed		Unemployed		Out-of-the-Labour-Force		Total
Employed	$u_{CJT} < \Delta t$	2221	17.1					2221
Unemployed	$u_{CJT} < \Delta t$	1774	13.6					1774
	$u_{CJT} \geq \Delta t$	1190	9.2					1190
	u_{CJT} missing	677	5.2					677
	$\Delta t > u_{ISU-2}$			551	100			551
						6606	100	6606
Out-of-the-Labour-Force	$u_{CJT} < \Delta t$	1770	13.6					1770
	$u_{CJT} \geq \Delta t$	3150	24.2					3150
	u_{CJT} missing	2221	17.1					2221
Total		13003	100%	551	100%	6606	100%	20160

Legend: u_{ISU-2} denotes an incomplete spell of unemployment in the second wave

3. Models for unemployment duration

Let T_i be a random variable denoting the completed unemployment spell of the *i*-th individual and t_i, its observed value in months or weeks or days. Let $\mathbf{x}_i(t)$ be a (row) vector of time-dependent and independent explanatory variables (covariates) representing the individual circumstances of the *i*-th job seeker such as, personnel characteristics (gender, age, education), family information (size, number of children, number of employed members) or labour market status (unemployment rate, benefits). A job seeker's conditional probability of leaving unemployment in a short interval $(t, t + dt)$ is given by the product of the two elements: the probability of receiving a job offer, $q[t, \mathbf{x}_i(t)]$, and the probability of accepting it, which implies that the wage offered exceeds the reservation wage, $x[t, \mathbf{x}_i(t)]$ (Lancaster, 1990; Narendranathan and Stewart, 1993). Therefore, the conditional probability of leaving unemployment within the short interval $(t, t + dt)$ is the hazard

$$h[t, \mathbf{x}_i(t)]\,dt = q[t, \mathbf{x}_i(t)]\left[\int_{x[t,\,\mathbf{x}_i(t)]}^{\infty} f(w)\,dw\right] = q[t, \mathbf{x}_i(t)]\,a\{x[t, \mathbf{x}_i(t)]\}, \qquad (1)$$

where $f(w)$ is the distribution of wages relative to the vacancies in the labour market. The probability $q[t, \mathbf{x}_i(t)]$ expresses the job seeker's possibilities of finding a vacancy and being offered the job, given that he or she is available. The acceptance probability, $a\{x[t, \mathbf{x}_i(t)]\}$, is generally affected by individual factors concerning the stopping strategy and the parameters of the vacancy-wage distribution.

It is possible to model the hazard function alone, ignoring the underlying choice structure, although it loses some distinctive aspects of the job seekers' behaviour. Different specifications of $h[t, \mathbf{x}_i(t)]$ allow the model to deal with the problem of nonstationarity, as well as with other features that may arise in empirical contexts (Amemiya, 1985, pp. 433-455). The most frequently reduced form is expressed in a proportional hazards model (Cox and Oakes, 1984)

$$h(t_i, \mathbf{x}_i) = \frac{f(t_i, \mathbf{x}_i)}{1 - F(t_i, \mathbf{x}_i)} = h_0[t_i; \mathbf{?}]\, j\,(\mathbf{x}_i; \mathbf{ß}), \qquad (2)$$

where $f(t_i, \mathbf{x}_i)$ and $F(t_i, \mathbf{x}_i)$ are the density and the distribution functions of T_i, respectively, $h_0(t_i, \mathbf{?})$ is the baseline hazard at time t_i for the *i*-th individual with $\mathbf{x}_i = 0$. The function $j\,(\mathbf{x}_i; \mathbf{ß})$ denotes the impact of covariates $\mathbf{x}_i$ on the baseline hazard, and $\mathbf{ß}$ is the vector of parameters associated with them. Various specifications of $h_0(t_i, \mathbf{?})$ account for different distributions of T_i and therefore, the shapes of the baseline hazard such as the exponential, which is memoryless because it implies a constant hazard, the Weibull, the gamma, the log-logistic, and so on (Kiefer, 1988; Lancaster, 1979). The log-linear form for $j\,(\mathbf{x}_i; \mathbf{ß}) = \exp(\mathbf{x}_i \mathbf{ß})$ is often used to express the impact of covariates (Cox, 1972) and the hazard function becomes $h(t_i, \mathbf{x}_i) = h_0(t_i, \mathbf{?}) \exp(\mathbf{x}_i \mathbf{ß})$.

Changes occurring over time in the characteristics of individuals and the effects of the initial attributes on the evolution of the professional positions/conditions, or on the transitions between labour force states, represent the main topics of unemployment duration analysis. One aspect concerns the attenuation or accentuation of the impact produced by the starting attributes throughout the subsequent experiences over time, *i.e.*, *state dependence* (Heckman, 1981). Another aspect concerns the completeness of the selected individual characteristics and/or their persistence over time, *i.e.*, some variations across the sampling units (or over time for the same sampling units) may not have been included in the model due to measurement errors or omission (*heterogeneity*). Although this aspect is intrinsic in all statistical models, it leads to peculiar effects in the analysis of unemployment spells (Blossfeld and Hamerle, 1992).

In a duration model, as formulated in equation (2), state dependence is expressed by the baseline hazard $h_0(t_i, \boldsymbol{\theta})$, while the heterogeneity is captured by a set of explanatory variables, $\mathbf{x}_i$. *Unobserved* heterogeneity induced by omitted variables may be introduced on the scale parameter of the distribution assumed for T_i, considered as a continuous random variable (Heckman and Singer, 1984a,b; Heckman, 1991). However, identification of the distribution for the scale parameter is not easy and the choice is conditioned by computational difficulties.

For the available data, T_i was assumed to be Weibull-distributed. Therefore, $h_0(t_i, \boldsymbol{\theta}) = \alpha\ \lambda\ (\lambda\ t)^{\alpha-1}$ and the hazard is monotonic: increasing if $\alpha > 1$, decreasing if $\alpha < 1$, and constant if $\alpha = 1$. Thus, a value of α greater (or lower) than 1 implies an increase (or decrease) in the conditional probability that an unemployed individual will find a job shortly, as the unemployment spell increases in length. A value of $\alpha = 1$ implies a memoryless job-seeking process, *i.e.*, the hazard function is constant and there is no duration dependence. The density function of T_i will be

$$f(t_i, \mathbf{x}_i;\ \lambda, \alpha, \boldsymbol{\beta}) = \alpha\ \lambda\ (\lambda\ t)^{\alpha-1} \exp\left[\mathbf{x}\boldsymbol{\beta} - (\lambda\ t)^{\alpha} \exp(\mathbf{x}\boldsymbol{\beta})\right], \qquad (3)$$

but the assumption that T_i is Weibull-distributed, $W(t;\ \lambda, \alpha)$, is data-dependent, *i.e.*, it will be ascertained by examining the data. If heterogeneity is introduced on the scale parameter, λ, it becomes a continuous random variable, Λ. An interesting solution is obtained by assuming again that Λ is Weibull-distributed, $W(\lambda; \vartheta, \alpha)$, with the scale parameter ϑ and the same shape parameter, α, of the distribution of T_i (Lancaster, 1990, pp. 65-70; Lalla and Pattarin, 2001). Let $f(\lambda; \vartheta, \alpha)$ be the density function of Λ. Then, the density of the distribution of completed spell lengths is given by

$$y\ (t_i, \mathbf{x}_i; J, a, \text{ß}) = \int_0^\infty f(t_i, \mathbf{x}_i;\ l, a, \text{ß})\, f(l; J, a)\, dl = \frac{a\, t_i^{a-1}\, J^a\, \exp(-\mathbf{x}\text{ß})}{\left[t_i^a + J^a \exp(-\mathbf{x}\text{ß})\right]^2} \qquad (4)$$

and its corresponding hazard function (Lalla and Pattarin, 2001) becomes

$$h_e\,(t_i, \mathbf{x}_i; J, a, \text{ß}) = \frac{a\, t_i^{a-1}}{t_i^a + J^a \exp(-\mathbf{x}\text{ß})}. \qquad (5)$$

The inclusion of unobserved heterogeneity, according to the previous procedure, generates a distribution that is more flexible than the Weibull because it has a hazard with a non-monotonic trend: it is equal to 0 for t_i, increases to a maximum, then decreases and $h_e\,(t_i, \mathbf{x}_i) \to 0$ for $t_i \to \infty$. The result seems to be satisfactory because the distribution has a flexible and attractive hazard shape, but it should be noted that the choice of the Weibull is arbitrary and the mean and variance are not defined for some values of a .

3.1. Telescoping effect

The measurement of unemployment duration for the *i*-th individual, t_i, is strictly based on his/her capacity to evaluate exactly the time interval spent seeking job or to remember the dates determining that spell. However, memory fades and becomes imprecise over time, generating two kinds of errors. One is leaving out an episode: he/she forgets a short period of employment. Another is expansion/contraction of time (telescoping), where an episode is remembered as occurring later or more recently than it actually did. The telescoping effect for unemployment spells would generate pronounced spikes on their empirical distribution (*heaping effect*), with respect to the expected smoothness of the frequency polygon. In fact, the true unobserved spell length of the *i*-th individual, t_i, could have been approximated to a preferred value because it is easier to communicate, t_{Hij}, such as a quarter, semester, year and their multiples (Sudman and Bradburn, 1973; Sikkel, 1985), *i.e.*, a sort of calendar effect.

The location of the spikes (*heaped values*) is a matter of empirical evidence. Let J be the number of heaped values and let $h_1 < h_2 < \cdots < h_J$ be their corresponding observed abscissas (one for each spike), respectively, sorted in ascending order. Let y_i be a binary variable: $y_i = 1$ with a probability of $H(t_i; \mathbf{?})$ for the observed spell length coinciding with a heaped value; otherwise $y_i = 0$ with a probability of $[1 - H(t_i; \mathbf{?})]$. Therefore, $H(t_i; \mathbf{?})$ is the *heaping* *function* with a vector of parameters **?**, denoting the probability of having detected a spell length, t_i, arising from the telescoping effect. A possible model for the stochastic heaping processes becomes $t_{Hij} = t_i + \boldsymbol{d}_{ij}\, y_i$, where $\boldsymbol{d}_{ij} = h_j - t_i$ is such that $\left| h_j - t_i \right|$ represents the minimum for $j = 1,2,\ldots,J$ (Torelli and Trivellato, 1993). The exponential distribution, $H(t_i; \boldsymbol{?}) = 1 - \exp(g)$, is the simplest and most practical assumption, although **i** may depend on individual characteristics, $\mathbf{x}_i$, which were not considered here in order to avoid over-parameterisation.

The binary variable y_i was defined for each type of spell: $y_{A;i}$ for $T_{A;i}$ and $y_{R;i}$ for $T_{R;i}$. Therefore, two sets of heaped values will be obtained. Let only $T_{A;i}$ be considered. The abscissas of heaped values, $h_{A;1} < h_{A;2} < \cdots < h_{A;J}$, were determined from the frequency distribution of $t_{A;i}$ (Fig. 2) and $T_{R;i}$ (Fig. 3). Now, it is possible to identify an interval for each $t_{A;i} = h_{A;j}$, considering two adjacent points $h_{A;j-1}$ and $h_{A;j+1}$. The lower bound is $t_{A;HLij} = (h_{A;j-1} + h_{A;j})/2$, and the upper bound is $t_{A;HUij} = (h_{A;j} + h_{A;j+1})/2$. Moreover, in general, a spell may be still in progress. The binary variable c_i could represent censored spells: it is equal to 1 if the observed unemployment spell t_i is complete and 0 otherwise. Therefore, two dummies are needed for the two types of spells, respectively $c_{A;i}$ for $T_{A;i}$ and $c_{R;i}$ for $T_{R;i}$. For the *i*-th individual, the available data set is given by $\{t_{A;i}, c_{A;i}, y_{A;i}, t_{A;HL\,ij}, t_{A;HU\,ij}, t_{R;i}, c_{R;i}, y_{R;i}, t_{R;HL\,ij}, t_{R;HU\,ij}, \mathbf{x}_i\}$, where $t_{\bullet;HL\,ij}$ and $t_{\bullet;HU\,ij}$ assume their corresponding values only when $y_{\bullet;i}$ is equal to 1 or 0 otherwise.

The empirical percentage distribution of $T_{A;i}$ shows many spikes, but their abscissas $\{3, 5, 13, 14, 19, 20, 25, 26, \cdots\}$ are not immediately interpretable in ordinary calendar time (Fig. 2), as their end points were calculated with a certain precision. On the contrary, the distribution of u_{ISU}, collected through Q46 and not reported here, showed more marked and recognizable abscissas values as years or submultiples like quarters and semesters: $\{3, 6, 12, 15, 18, 24, 36, 48, \cdots\}$. The empirical percentage distribution of $T_{R;i}$ in Fig 3 shows few marked spikes, as the dates are not subject to the calendar effect on the part of respondents. Therefore, the distribution of $T_{R;i}$ is smoother than that of $T_{A;i}$, except in some points: $\{4, 8, 14, \cdots\}$.

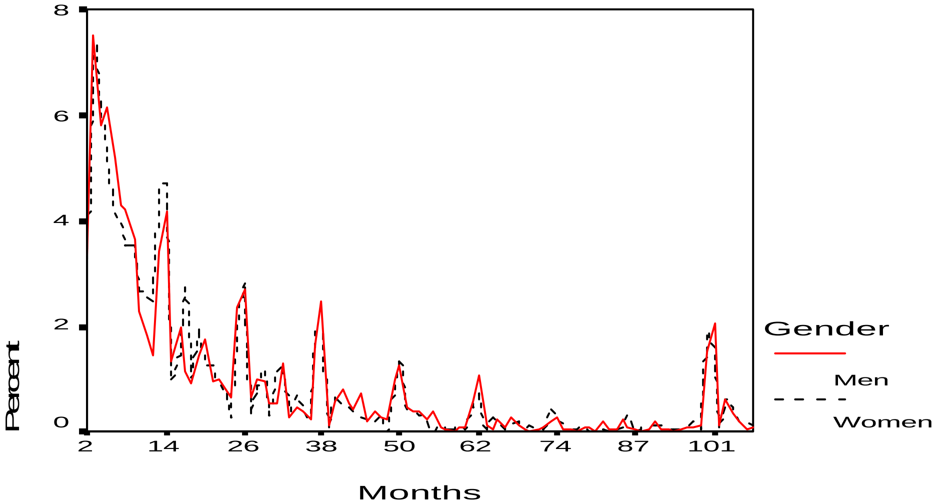

Fig. 2 – Percentage distributions of spell answers, $T_{A;i}$, for men and women (1997:4-1999:1).

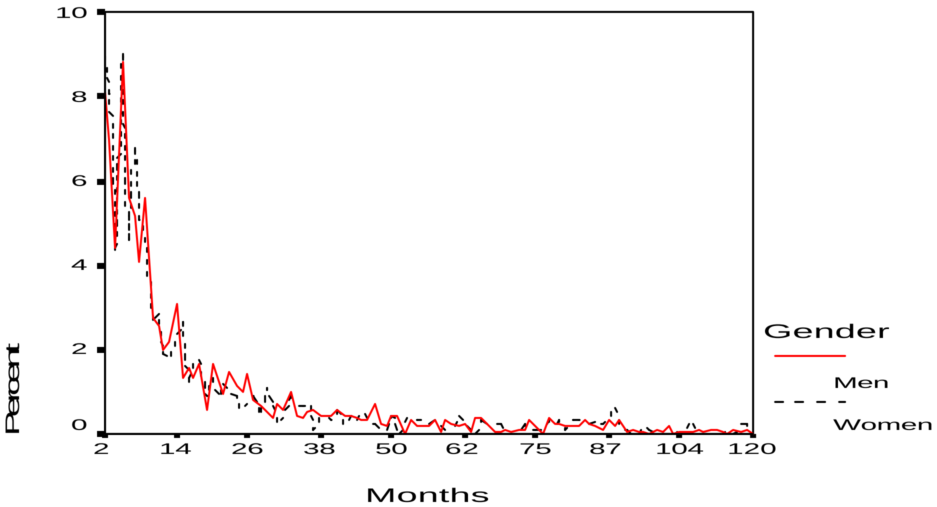

Fig. 3 – Percentage distributions of spell reconstructions, $T_{R;i}$, for men and women (1997:4- 1999:1).

3.2. Mixture models and parameter estimations

Let T_i be the unobserved completed spell of the *i*-th individual, considered as a random variable. The values of T_i concern the observation of the random variables $T_{A;\,i}$ and $T_{R;\,i}$, *i.e.*, spell answers and spell reconstructions, respectively. Therefore, the distribution of T_i could be expressed by a mixture distribution of $T_{A;\,i}$ and $T_{R;\,i}$:

$$f(t_i, \mathbf{x}_i; \mathbf{?}, \mathbf{ß}) = p\; y_{A}(t_{A;\,i}, \mathbf{x}_i; \mathbf{?}, \mathbf{ß}) + (1-p)\, y_{R}(t_{R;\,i}, \mathbf{x}_i; \mathbf{?}, \mathbf{ß}), \tag{6}$$

where p is the weight of the linear combination of the two distribution functions, $y_{A}(t_{A;\,i}, \mathbf{x}_i; \mathbf{?}, \mathbf{ß})$ and $y_{R}(t_{R;\,i}, \mathbf{x}_i; \mathbf{?}, \mathbf{ß})$ expressed in equation (4), for $T_{A;\,i}$ and $T_{R;\,i}$, respectively. The general form of the functions in the linear combination would include the heaping effect and the right-censored spells, which are most frequent in duration data. As an example, for the random variable $T_{A;\,i}$, the density function is

$$\begin{aligned} y^{*}_{A}(\cdot) = {} & \Big\{ \big[y_{A}(t_{A;\,i}, \mathbf{x}_i; \mathbf{?}, \mathbf{ß})\big]^{c_i} \big[S_A(t_{A;\,i}, \mathbf{x}_i; \mathbf{?}, \mathbf{ß})\big]^{1-c_i} \big[1 - H_A(t_{A;\,i}; \mathbf{?})\big] \Big\}^{y_i} \\ & \times \Big\{ \big[y_{A}(t_{A;\,i}, \mathbf{x}_i; \mathbf{?}, \mathbf{ß})\big]^{c_i} \big[S_A(t_{A;\,i}, \mathbf{x}_i; \mathbf{?}, \mathbf{ß})\big]^{1-c_i} \big[1 - H_A(t_{A;\,i}; \mathbf{?})\big]^{c_i} \\ & + \int_{t_{HLij}}^{t_{HUij}} \big[y_{A}(v, \mathbf{x}_i; \mathbf{?}, \mathbf{ß})\big]^{c_i} \big[S_A(v, \mathbf{x}_i; \mathbf{?}, \mathbf{ß})\big]^{1-c_i} \big[H_A(v; \mathbf{?})\big]\, dv \Big\}^{1-y_i}, \end{aligned} \tag{7}$$

where $S_A(\cdot)$ is the corresponding survivor function, $H_A(\cdot)$ is the heaping function, y_i is the binary variable denoting whether $t_i = h_j$, and c_i is again a binary variable denoting whether the observed spell is a completed spell. An explanation of this expression can be found in Lalla and Pattarin (2001).

The transitions considered in the following only generate completed spells and c_i is always equal to 1. If the heaping effect is not considered, the log-likelihood function is then given by

$$\begin{aligned} \log L(p, J, a, \mathbf{ß}) = {} & n \log(a) + n\, a \log(J) - \sum_{i=1}^{n} \left(\sum_{j=1}^{m} X_{ij}\, b_j \right) \\ & + \sum_{i=1}^{n} \log\big[p\, \tilde{y}(t_{R;\,i}, \mathbf{x}_i; J, \mathbf{ß}) + (1-p)\, \tilde{y}(t_{A;\,i}, \mathbf{x}_i; J, \mathbf{ß})\big]. \end{aligned} \tag{8}$$

where n is the total number of observations, m is the total number of covariates, $\tilde{y}(\cdot)$ is the density function of T_i (which could be $T_{A;\,i}$ or $T_{R;\,i}$), a is the shape parameter, q is the location parameter, and $\mathbf{ß}$ is the usual vector of coefficients for the covariates. The density functions $\tilde{y}(\cdot)$ are derived from equation (4) and assume the form

$$\tilde{y}(t_i, \mathbf{x}_i; J, a, \mathbf{ß}) = \frac{t_i^{a-1}}{\left[t_i^{a} + J^{a} \exp(-\sum_{j=1}^{m} X_{ij} b_j) \right]^2}. \tag{9}$$

Inclusion of the heaping effect highly complicates the expression of $\tilde{y}(\cdot)$, as can be deduced from equation (7).

4. Differential Evolution: some background information

Evolutionary algorithms are stochastic search heuristic algorithms. They can deal with complex optimisation problems. An evolutionary algorithm starts by randomly generating a population of individuals, which encode feasible candidate solutions to the optimisation problem under investigation. Each individual is the mathematical encoding of one candidate solution to the optimisation problem under investigation. A fitness function, which measures the goodness of a candidate solution, is then computed for each individual in the population. The population is then altered, by iteratively applying some mathematical operators. This iterative process is usually defined as the evolution of the population through generations. For each generation some mathematical operators, namely selection, recombination and mutation, are applied to the individuals of the current population in order to determine the composition of the population of the next generation. Such operators aim to determine candidate solutions with a better fitness value by exploring the search space. The population is evolved through different generations until the algorithm converges to a satisfying solution or the maximum number of iterations is reached. The fittest individual in the final population then encodes the optimal solution to the problem under investigation.

Genetic algorithms (Holland, 1975) are probably the most well-known evolutionary algorithms, even outside the heuristic community. They have been successful in tackling optimisation problems in different fields, such as statistics, engineering, and bioinformatics. Recently, many empirical investigations have shown that Differential Evolution (DE), a relatively unknown approach to numerical optimization (Storn and Price, 1995), has a superior performance compared to other optimisation and evolutionary algorithms. DE is very simple to implement and requires little or no parameter tuning.

DE starts by randomly generating an initial population (P_{Gen}) of N strings $\boldsymbol{x}_i$ $(i = 1, \ldots, N)$, which are the individuals in the population. Each individual (chromosome) consists of a string of D real-value cells (genes), which encodes a candidate solution (genome) to the optimisation problem under investigation. After random initialisation and fitness evaluation, the population is then evolved through generations ($Gen = 1,\ldots,G_{max}$). In each generation, mutation, recombination and selection operators are applied to the individuals. In every generation, for each individual i, we choose three other individuals k, l, and m randomly from the population (with $j^{1}k^{1}l^{1}m$), calculate the difference of the individual in k and l, scale it by multiplication with a parameter F and create an offspring by adding the result to the chromosome of m. However, the entire individual of the offspring is not created in this way, but genes are partly inherited from individual i. The proportion of genes that are also partly inherited is determined by the so-called crossover probability (CR), which determines how many genes of the difference vector are copied on the average in the offspring. More precisely, the operator iteratively copies consecutive genes (from a random starting point and continuing on with the first gene after the last gene has past) of the difference vector in the offspring until $U(0,1) \geq CR$. If the offspring o is better than the parent i, then i is substituted by o.

The process is repeated for a fixed number of iterations and the optimization result is the best recorded candidate solution and fitness value at the end of the run. Different stopping criteria can be used. Table 4 reports the pseudo-code of the differential evolution (DE) algorithm.

4.1. Differential evolution encoding for the unemployment spells mixture

The DE algorithm could be useful in estimating the parameters of the mixture distribution of the unemployment spells when the spell reconstructions ($T_{R;\,i}$) and spell answers ($T_{A;\,i}$) are considered jointly in a mixture model. The aim is to determine the parameters that maximise the likelihood function of the mixture model or minimise the negative of the log-likelihood function (see equation 8)

The analysis was performed considering 2380 individuals, from a sample with no missing values, collected by ISTAT, considering only the transitions from the unemployment to the employment state (see Table 3). For each person in the sample, m=51 covariates were considered. As Fig. 4 shows, each DE individual encodes the following parameters: the mixture weight p, the shape parameter a, the location parameter J, and the 51 coefficients of the covariates

$(b_1, \ldots, b_m)$, where m=nvar. The DE explores to determine the parameters that minimise the following fitness function, which corresponds to the log-likelihood of the model described by (8) and (9)

$$fitness = -\log L(p, J, a, ß)\,. \tag{10}$$

DE requires a priori setting of the search space domain and some evolutionary parameters. After some preliminary analyses, we set the population size at 165 (NIND=3*N.Parameters to be estimated), the maximum number of generations G_{Max} at 4000, the crossover rate at 0.5 (CR), and the DE-step size at 0.8 (F). Moreover, the search space was constrained so that the mixture weight p can vary in $[0,1]$, the shape parameter a in $[0,4]$, the location parameter J in $[0,8]$, and the 51 beta parameters $(b_1, \ldots, b_m)$ in $[-2, 2]$. During the simulations, a parameter was set to the closest domain boundary in case it proved to be equal to a value outside the domain search space. Future investigation will consider a larger search space.

Table 4
Pseudo-code of the DE algorithm

```
void differential_evolution()
{
  initialize();
  evaluate();
  While (Gen<Gmax; Gen++) {
    for (i=1; i<NIND; NIND++) {
      for (j=1; j<D; D++) {
      MutateandRecombine();
      }
    select();
    evaluate();
    }
  }
}
```

```
Void MutateandRecombine(){
```

$$o.\mathtt{gene_j} = \begin{cases} \mathtt{m.gene_j} + \mathtt{F}\cdot(\mathtt{k.gene_j} - \mathtt{l.gene_j}) & \mathtt{if\ U(0,1)} < \mathtt{CR} \\ \mathtt{i.gene_j} & \mathtt{otherwise} \end{cases}$$

```
}
```

```
Void select(){
  If fitness(offspring)>fitness(parent), choose offspring;
  Otherwise, choose parent;
}
```

5. Results of the estimation parameters

The data set was restricted solely to transitions from unemployment to employment, considering only job losers/leavers. The averages of the two durations, $t_{A;i}$ and $t_{R;i}$, were almost the same (see Table 5), but values of $t_{R;i}$ greater than 120 months were eliminated (see §2). In any case, this indicates that the two durations were consistent, on the average, while the mean of their minimums showed a reduction of five months. The means of $t_{A;i}$ and $t_{R;i}$ were somewhat higher than those obtained by Favro-Paris et al. (1996), while the mean of min($t_{A;i}$; $t_{R;i}$) was slightly lower than their value. Only $t_{R;i}$ proved to have missing values for the unemployed and they amounted to 206 cases. To avoid the elimination of these statistical units, the missing values of $t_{R;i}$ were replaced with the corresponding available value of $t_{A;i}$, given that their means were almost the same.

The descriptive statistics of the main variables used to model unemployment durations are reported in Table 5, grouped by personal, family, individual job-search, local job-search, previous job, and current job characteristics. The explanatory variables were selected using two main criteria, although they admittedly remain arbitrary. The first used the backward method and maximisation of the partial likelihood estimates for removal, in Cox's semi-parametric proportional hazard model for $t_{A;i}$, for $t_{R;i}$, and for their minimum. The final set was obtained by matching the three outcomes. The second used both personal judgements and the main findings in the literature to select covariates from among those that did not enter the models.

The personal characteristics included gender, age, and education. The proportion of women involved in these types of transitions was only 40.5%, suggesting discrimination (and segregation) in the labour market. In fact, the number of unemployed women was generally greater than that of men, but only a lower proportion of them proved to have found employment and many of them passed into the out-of-the-labour-force state. Age was calculated at the beginning of unemployment and expressed in decades to reduce its value with respect to the dummies. It was introduced in the models as a polynomial of the second order to account for the different behaviours of individuals in different life phases. The level of education did not reveal a significant impact on the transition from unemployment to the employment state and it was not included in the model, for the sake of brevity, even if in other studies, they have sometimes yielded coefficients significantly different from zero (see, *inter alia*, Røed and Nordberg, 2003; Böheim and Taylor, 2002).

Table 5
Acronym (Acr.), Mean, Standard Deviation (SD), Range, and number of cases (n) –or 1 when the variables is binary (0/1)– for the main variables in the models

Variable	Acr.	Unemployed Mean	SD	Range	n or no. of 1
Spell Answers, $t_{A;i}$ (months)		19.760	22.889	1.9–107	2585
Spell Reconstructions, $t_{R;i}$ (months)		19.352	22.952	1.5–119	2379
Min($t_{A;i}$; $t_{R;i}$) – months		15.174	17.770	1.5–107	2585
Personal characteristics					
Women	W	0.405	0.491	0/1	1046
Age (in decades)	A	3.276	1.069	1.1–8.0	2585
High School (*Diploma*)	HS	0.251	0.433	0/1	648
College Level (*Laurea*)	CL	0.056	0.229	0/1	144
Family characteristics					
Never Married	NMarr.	0.471	0.499	0/1	1217
Married	Marr.	0.489	0.500	0/1	1264
Separated, Divorced, Widowed	SDW	0.040	0.197	0/1	104
Child (compared to the head of family)	Child	0.418	0.493	0/1	1081
Number of Family Components	NFC	3.637	1.300	1–12	2585
Employment Ratio in the Family	ERF	0.191	0.204	0–0.75	2585
Individual job-search characteristics					
Short-Term Contract	STC	0.041	0.197	0/1	105
Part-Time Work	PTW	0.102	0.303	0/1	264
Any Time	AT	0.236	0.425	0/1	610
Not Willing to Commute	NWC	0.331	0.471	0/1	856
Willing to Commute	WC	0.459	0.498	0/1	1187
Any Location	AL	0.210	0.407	0/1	542
Conditional Availability to Work	CAW	0.621	0.485	0/1	1604
Number of Job Search Attempts	JSA	1.974	1.365	0–10	2585
Reservation Wage	RW	1.694	1.589	1–31.5	2585
Registered with Employment Agency	REA	0.832	0.374	0/1	2150
Benefits	Bnft	0.164	0.370	0/1	424
Local job-search characteristics					
Provincial Unemployment Rate	PUR	15.028	8.724	2.2–34.4	2585
North-East (4, 5, 6)	NE	0.113	0.317	0/1	293
Centre: Emilia-Romagna	CE-ER	0.060	0.237	0/1	154
Centre (9, 10, 11)	CE	0.098	0.297	0/1	253
Latium	LZ	0.051	0.219	0/1	131
South-East (13, 14, 16)	SE	0.138	0.345	0/1	357
1st Quarter	D1Q	0.234	0.424	0/1	606
2nd Quarter	D2Q	0.289	0.453	0/1	747
3rd Quarter	D3Q	0.133	0.340	0/1	345
4th Quarter of previous year	D4Qpy	0.220	0.414	0/1	568

(continued on the next page)

Table 5 (*continued*)

Variable	Acr.	Unemployed Mean	SD	Range no. of 1	*n* or no. of 1
Previous job characteristics					
Dismissed from the Previous Job	DPJ	0.240	0.427	0/1	621
Left for Other Reasons	LOR	0.203	0.403	0/1	526
Agriculture Sector	AS1	0.123	0.329	0/1	318
Construction Sector	CS1	0.175	0.380	0/1	453
Trade Sector	TS1	0.220	0.414	0/1	569
Service Sector	SS1	0.101	0.302	0/1	262
Public Administration	PA1	0.115	0.319	0/1	297
Current job characteristics					
Working in a Different Region	WDR	0.474	0.499	0/1	1226
Forced Part-Time	FPT	0.147	0.354	0/1	380
Chosen Part-Time	CPT	0.063	0.242	0/1	162
Forced Temporary Employment	FTE	0.271	0.444	0/1	700
Chosen Temporary Employment	CTE	0.172	0.377	0/1	444
Owner Position	OP	0.042	0.201	0/1	109
Professional Man	PM	0.109	0.312	0/1	283
Coadjuvant Position	CP	0.026	0.160	0/1	68
Agriculture Sector	AS2	0.159	0.366	0/1	412
Construction Sector	CS2	0.168	0.374	0/1	435
Trade Sector	TS2	0.261	0.439	0/1	674
Service Sector	SS2	0.138	0.345	0/1	356
Public Administration	PA2	0.113	0.317	0/1	293

Family characteristics were limited to marital status, the «position» in the family, the number of family components, and the employment ratio in the family. Marital status was subdivided in three dummies: never married, married, and other (separated, divorced, widowed). The position in the family contains dummies distinguishing between head of family, spouse, child, and other. The employment ratio is the division of the number of employed by the number of family components, ranging from 0 to 1.

The individual job-search characteristics also included three covariates that were not significant. Conditional availability to work is a negative attitude with respect to unconditional availability, the latter being a preliminary condition to be classified as unemployed. The number of job search attempts represents an indicator of the intensity of the job-search, but a high value could denote some intrinsic problems of job seekers (Favro-Paris et al., 1996). The reservation wage is an important reference for job search theory (Lippman and McCall, 1976), but it has not proven to be significant in other studies (Lalla and Pattarin, 2001). However, it is known that answers regarding wages are generally missing and often unreliable, introducing a bias in the estimates.

The local job-search characteristics include the Provincial unemployment rate, which was constant over the period considered. This rate captures local differences without fulfilling them. Therefore, some dummies for the geographic areas of Italy were introduced in the model using the aggregations proposed by Attanasio and Padoa Schioppa (1991), with a small change in the definition of Centre where the region of Emilia-Romagna was introduced as a single area (see Table 5). Other dummies concern the temporal period (i.e., quarter) before finding a job: the reference dummy was the fourth quarter of 1998.

All the available previous job characteristics were included in the model and they proved to have a significant impact on the probability of leaving the unemployment state, while current job characteristics did not always yield coefficients significantly different from zero. In particular, the sectors proved to be significant only in the model based on $t_{R;i}$. However, the latter were included for completeness.

Table 6 reports the estimates of the parameters of the spell mixture models when 15 runs of DE were considered. In each run, the population was iteratively evolved for 4000 generations. Fig. 4 shows that (after 500 generations) the best fitness value in each generation decreased markedly at the beginning (0-500 generations) and then (after 500 generations) the fitness value of the best individual continued to decrease, but in a more stable manner. Four thousand generations can be acceptable even if a better solution to the optimisation problem is found when the number of generations increases.[1] Column 2 shows the «best» values of b, column 3 the mean values, and column 4 the standard error of the parameters for 15 runs. The other columns show, respectively, the value of b, its standard error, the relative p-value ($p <$), and the corresponding estimates of the percentage of change in risk with each unit change in the covariate or exp($ß_i$), which can be interpreted as the relative risk when the covariate takes only the values of 0 and 1. Cox's model was based only on the spells given by min($t_{A;i}, t_{R;i}$) and the parameters l and a were estimated through a nonlinear model of the empirical baseline cumulative hazard. The coefficient of determination for this post-hoc estimation was $R^2 = 0.992$.

The estimated value, $\hat{a}$, of the shape parameter is particularly interesting because some labour market theories predict $\hat{a} < 1$, implying negative state dependence (Salant, 1977), but the empirical results showed all possible values, i.e., either negative (Ham and Rea, 1987; Torelli and Trivellato, 1988; Lynch, 1989), or positive (Hujer and Sneider, 1989), or null (Flinn, 1986; Groot, 1990)

[1] All experiments were run using Windows XP on an IBM desktop PC. With the parameters described above, each DE simulation required 8h 26m. The estimation of the parameters in Cox's regression model was obtained through SPSS (1997).

state dependence, although the segment of the surveyed population was not always the same. The estimated shape, $\hat{a}$, showed a value slightly, but significantly, lower than 1, which could depend on the noticeable heterogeneity not being captured completely by the selected covariates. The estimates of J (or l) refer to the average spell length, as it is the scale parameter. The $\hat{b}$ coefficient represents the predicted change for a unit increase in the predictor. Therefore, a positive value implies an increase in the probability (acceleration) of leaving the unemployment state with the exp($ß_i$) being greater than 1, while a negative value entails a decrease in the probability (deceleration) of leaving the unemployment state with the exp($ß_i$) being lower than 1.

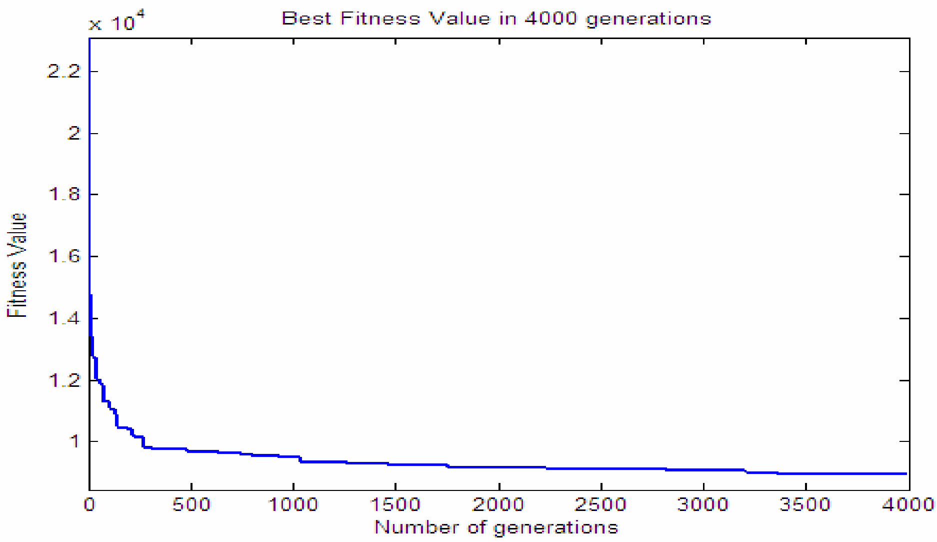

Fig. 4. Fitness value of the DE optimal solution in 4000 generations.

Of the personal and family characteristics, only the “child” covariate proved to have a positive coefficient $\hat{b}$ statistically different from zero, implying a reduction of unemployment duration. Children looking for a job could be induced to accept an offer because either the family could be pressed for money or the employed in the family could represent an aid in finding a job, but a negative sign also has an interesting interpretation. Therefore, its sign is not univocally determined *a priori*.

Among the individual job-search characteristics, only the covariates REA and “Benefits” proved to be significant. The coefficient of REA was negative, as expected, because it denotes a deliberate intention to find a job. A positive coefficient for benefits does not have a causal interpretation, as it could represent a spurious correlation with the omission of important firm-specific variables from the model (Røed and Nordberg, 2003). However, it is often argued that recipients should wait longer than those without benefits, but the empirical outcomes sometimes contradict one another (see, inter alia, Nickell, 1979b; Narendranathan et al., 1985). In this case, benefits tend to increase unemployment duration.

The significant local job characteristics included the Provincial unemployment rate with an expected negative coefficient and the dummies for the North-East with an expected positive coefficient. The other dummies ? Latium, the South-East, the second quarter, and the fourth quarter of the previous year? yielded negative coefficients. The second and the fourth quarters denote some seasonal effects on unemployment flows, as they indicate periods in which job offers are high for seasonal workers.

All the available previous job characteristics proved to be significant and the coefficients showed the expected signs. Dismissed workers and individuals belonging to the residual category of "leavers for other reasons" showed longer unemployment lengths than those of voluntary leavers. Workers in industry proved to have spells longer than those of workers in agriculture, construction, trade, services, and also even public administration. For agriculture, construction, and the trade sector it could depend on the presence of a large number of seasonal workers, while the service sector and public administration could denote the specificity of the current labour market in which industry tends to be less important in the dynamics of unemployment.

The selected current job characteristics were all significant, except the dummies for the sectors. Individuals accepting (forced or chosen) temporary employment showed an increase in the probability of leaving the unemployment state. Owner positions, professionals and coadiuvant positions yielded negative coefficients with respect to the other positions, such as clerical, low- and high-skilled workers.

Estimates of parameters, obtained through the DE algorithm, concerned the values of the "best" $\hat{b}$ coefficients, i.e., the values of parameters generating the lowest value of the likelihood. Each run produces a set of values that is the best in that population. Therefore, the mean of the various "best values" of the same $\hat{b}$ coefficient, obtained in the 15 runs, is also reported. Finally, the evaluation of the significance of the parameters required the calculation of the corresponding standard errors. The naïf procedure adopted estimated the standard error of each parameter using the standard deviation of the different best values obtained in the various runs. The results are reported in Table 6.

Table 6
Estimates of parameters of the mixture model for completed spells of unemployment obtained by the DE method —the “best” b, the mean of the $\bar{b}$ s, and the relative standard error SE(b)— and for Cox’s proportional hazards model, using min($t_{A;i}$; $t_{R;i}$)

Variables	DE estimates			Cox’s proportional hazards model			
	Best b	$\bar{b}$	SE(b)	b	SE(b)	$p <$	exp(b)
$\hat{p}$	**0,276	**0,267	0,053	—	—	—	
$\hat{a}$	**2,133	**2,094	0,092	0.886	0.001	**0.001	
$\hat{J}$ / $\hat{l}$	**6,254	**6,472	1,752	0.083	0.003	**0.001	
g	0,000	0,000	0,000				
W	0,514	–0,026	0,281	0,032	0,049	0,510	1,033
A/10	0,334	0,333	0,336	0,050	0,125	0,686	1,052
(A/10)^2	–0,039	–0,028	0,050	0,009	0,016	0,597	1,009
Child	0,155	0,317	0,323	0,196	0,056	**0,000	1,217
NCF	0,089	0,096	0,097	0,040	0,016	*0,013	1,041
ERF	0,077	0,168	0,519	0,122	0,108	0,261	1,130
STC	–0,077	0,416	0,532	0,217	0,104	*0,037	1,242
PTW	–0,471	–0,285	0,335	–0,115	0,073	0,114	0,891
AT	–0,322	–0,126	0,239	–0,087	0,050	0,079	0,916
AL	*0,529	0,146	0,250	0,067	0,051	0,190	1,070
CAW	*0,441	0,107	0,215	0,043	0,044	0,330	1,044
JSA	*–0,160	–0,056	0,078	0,013	0,016	0,428	1,013
RW	0,031	–0,024	0,071	0,007	0,012	0,535	1,007
REA	*–0,706	*–0,604	0,267	–0,255	0,057	**0,000	0,775
Bnft	*1,080	*0,848	0,397	0,513	0,061	**0,000	1,670
PUR	*–0,036	**–0,051	0,020	–0,019	0,003	**0,000	0,982
NE	0,635	0,282	0,405	0,178	0,077	*0,022	1,194
CE-ER	0,299	–0,034	0,517	0,099	0,094	0,290	1,104
CE	–0,321	–0,441	0,330	–0,101	0,076	0,186	0,904
LZ	**–1,340	*–0,965	0,436	–0,449	0,096	**0,000	0,639
SE	–0,158	–0,083	0,351	–0,124	0,061	*0,041	0,884
D1Q	**–1,245	–0,592	0,421	–0,100	0,074	0,178	0,905
D2Q	**–1,223	*–0,835	0,370	–0,244	0,075	**0,001	0,784
D3Q	–0,005	0,153	0,471	0,074	0,079	0,349	1,077
D4Qpy	**–1,304	**–1,210	0,370	–0,371	0,086	**0,000	0,690
DPJ	–0,329	–0,315	0,288	–0,182	0,053	**0,001	0,834
LOR	–0,110	–0,367	0,286	–0,139	0,056	*0,013	0,870
AS1	**1,166	**1,357	0,492	0,648	0,086	**0,000	1,912
CS1	0,375	0,648	0,521	0,338	0,078	**0,000	1,403
TS1	*0,522	**0,722	0,279	0,356	0,063	**0,000	1,428
SS1	0,226	0,665	0,443	0,383	0,077	**0,000	1,467
PA1	**1,022	**1,222	0,422	0,532	0,077	**0,000	1,703

(continued on the next page)

Table 6 (*continued*)

Variables	DE estimates			Cox's regression model estimates			
	Best b	$\bar{b}$	SE(b)	b	SE(b)	$p <$	$e^{\bar{x}ß}$
WDR	**–0,708	**–0,854	0,223	**–0,275	0,053	0,000	0,759
FPT	–0,193	–0,320	0,344	**–0,241	0,060	0,000	0,786
CPT	–0,563	–0,260	0,565	*–0,189	0,086	0,028	0,827
FTE	0,270	0,125	0,194	0,100	0,054	0,065	1,105
CTE	–0,049	0,245	0,353	*0,155	0,060	0,010	1,168
OP	*–1,592	–0,669	0,708	*–0,256	0,110	0,020	0,774
PM	–0,533	–0,378	0,403	*–0,176	0,072	0,014	0,839
CP	–0,774	–0,690	0,663	**–0,477	0,131	0,000	0,621
AS2	–0,319	–0,351	0,403	–0,145	0,087	0,096	0,865
CS2	–0,497	–0,372	0,425	–0,107	0,085	0,210	0,899
TS2	–0,427	–0,380	0,297	–0,030	0,069	0,666	0,971
SS2	–0,514	–0,466	0,341	–0,118	0,078	0,130	0,888
PA2	–0,085	–0,492	0,473	–0,146	0,086	0,090	0,864
–log(L)	**8779.863**	**9825.392**					

[bl] Border-Line. * Significant at the 5% level. ** Significant at the 1% level.

The DE algorithm yielded the estimates of all parameters in the model. The estimate of p indicated that $t_{R;i}$ was heavier than $t_{A;i}$, while the estimate of g proved to be negligible. It should be noted that the standard errors of the estimates were higher than those obtained by the partial maximum likelihood applied in estimating the parameters of the Cox's proportional hazards model (Cox, 1972). Therefore, the adopted procedure should be refined or the BHHH method (Greene, 1993) should be used, i.e., the outer product of the gradient which requires only the calculation of the first derivatives. Moreover, it is not useful to compare the different significance of parameters, as the latter depend on the value of their corresponding standard errors and the models are not exactly the same. However, only two significant covariates presented different signs: the short-term contract (STC) parameter, which was significant at the 5% level, and the parameter of chosen temporary employment (CTE) with borderline significance (6%).

6. Conclusions

This paper has examined unemployment durations in Italy detected in the period ranging from 1997:4 to 1999:1. Two measures were determined: spell answers and spell reconstructions. As is known, results obtained from unemployment spells analyses alone are generally ambiguous. In fact, conclusions concerning policy indications are not grounded upon a valid and stable basis (Heckman and MaCurdy, 1992). However, the analysis of unemployment duration remains important for an approximate understanding of the pattern of relationships between covariates.

On the average, the spell reconstructions proved to be similar, to the spell answers, when the former were bounded within the same range as the latter. Therefore, the missing data for one type of measure could be (and were) replaced with the data of the other type.

The use of the mixture distribution to model unemployment spell lengths could constitute an alternative «data-based» strategy against the choice of the minimum between the two types of spells. The mixture parameter could be interpreted as the weight of one with respect to the other, but it tends to increase the average length of the spells.

The DE method proved to be a valuable alternative to classical methods, although it is computationally expensive and convergence can be slow. However, it appears to be very useful in dealing with the complex log-likelihood function because the optimisation problem is then NP-hard and classical methods cannot provide a solution. In fact, the latter require the determination of the first (and sometimes also the second) derivatives, a low number of covariates, and a rather simple type of function. On the contrary, although an unsolved problem still remains concerning the calculation of the standard errors and the procedure used above requires further refinements, the limitations of the other methods are not present in the DE method.

Acknowledgements

The work reported here was partially supported by a grant from the Italian Ministry for Education, University and Research (MIUR) for the project, «Industrial Districts as Complex Systems», approved in 2002 (prot. 2002133972) and directed by David Avra Lane. Partial financial support was also provided by the Italian MIUR for the project «Dynamics and inertia in the Italian labour market: databases, measurement and modelling issues», coordinated by Ugo Trivellato. Individual anonymous data, elicited from the Italian Labour Force Survey, were kindly provided by Istat (the Italian national statistical agency), under a research agreement with the Department of Statistics of the University of Padua.

References

Amemiya, T., 1985. *Advanced Econometrics*. Harvard University Press, Cambridge (MA).

Attanasio, O.P., Padoa Schioppa, F., 1991. Regional Inequalities, Migration and Mismatch in Italy, 1960-86. In: Padoa Schioppa, F. (Ed.), *Mismatch and labour mobility*. Cambridge University Press, Cambridge, pp. 237–320.

Blossfeld, H.-P., Hamerle, A., 1992. Unobserved heterogeneity in event history models. *Quality & Quantity* 26, 157–168.

Böheim, R., Taylor, M.P., 2002. The search for success: do the unemployed find stable employment?, *Labour Economics* 9, 717-735.

Commission Regulation (EC) No. 1897/2000 of 7 September 2000. Implementing Council Regulation (EC) No. 577/98 on the organisation of a labour force sample survey in the Community concerning the operational definition of unemployment, *Official Journal of the European Communities*, L228, Vol. 43, 8 September 2000, pp. 18–21.

Cox, D.R. 1972. Regression Models and Life Tables. *Journal of the Royal Statistical Society* B 34, 187–220.

Cox, D.R., Oakes, D., 1984. *Analysis of Survival Data*. Chapman and Hall, London.

Duncan, G.J., Kalton, G., 1987. Issues of Design and Analysis of Surveys Across Time. *International Statistical Review* 55 (1), 97–117.

EUROSTAT, 1996. *The European Union Labour Force Survey – Methods and Definitions*. EUROSTAT, Luxembourg.

EUROSTAT, 2002. *European Social Statistics – Labour Force Survey Results 2002*. EUROSTAT, Luxembourg.

Favro-Paris, M.M., Gennari, P., Oneto, G.P., 1996. La durata della disoccupazione in Italia: un'applicazione della struttura longitudinale dell'indagine sulle forze di lavoro. ISTAT, Quaderni di Ricerca, nuova serie 4: 1–79.

Flinn, J.C., 1986. Econometric Analysis of CPS-Type Unemployment Data. *The Journal of Human Resources* XXI, 456–484.

Greene, W.H. (1993). *Econometric Analysis*, 3rd edition. Macmillan, New York.

Groot, W., 1990. The Effects of Benefits and Duration Dependence on Re-Employment Probabilities. *Economics Letters* 32 (4), 371–376.

Ham, J.C., Rea, S.A. Jr., 1987. Unemployment Insurance and Male Unemployment Duration in Canada. *Journal of Labour Economics* 5 (3), 325–353.

Heckman, J.J., 1981. Heterogeneity and State Dependence. In: Sherwin Rosen (Ed.), *Studies in Labor Markets*. University of Chicago Press, Chicago, pp. 91–139.

Heckman, J.J., 1991. Identifying the Hand of Paste: Distinguishing State Dependence from Heterogeneity. *The American Economic Review* 81 (2), 75–79.

Heckman, J.J., Singer, B., 1982. The identification problem in econometric models for duration data. In: Hilderbrand, W. (Ed.), *Advances in Econometrics*. Cambridge University Press, Cambridge, pp. 39–77.

Heckman, J.J., Singer, B., 1984a. Econometric Duration Analysis. *Journal of Econometrics* 24, 63–132.

Heckman, J.J., Singer, B., 1984b. A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data. *Econometrica* 52, 271–320.

Heckman, J.J., MaCurdy, T., 1992. Empirical Tests of Labor Market Equilibrium: An Evaluation. In: Darity, W. Jr. (Ed.), *Labor Economics: Problems in Analyzing Labor Markets*. Kluwer Academic, Dordrecht, pp. 117–139.

Holland, J.H., 1975. *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Harbor.

Hujer, R., Schneider, H., 1989. The Analysis of Labor Market Mobility Using Panel Data. *European Economic Review* 33 (2/3), 530–536.

Kalbfleisch, J.D., Prentice, R.L., 1980. *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, New York.

Kiefer, N.M., 1988. Economic Duration Data and Hazard Functions. *Journal of Economic Literature* XXVI, 646–679.

Kiefer, N.M., Lundberg, S.J., Neumann, G.R., 1985. How Long Is a Spell of Unemployment? Illusions and Biases in the Use of CPS Data. *Journal of Business & Economic Statistics* 3 (2), 118–128.

Kleinbaum, D.G., 1996. *Survival Analysis: A Self-Learning Text*. Springer, New York.

Lalla, M., Pattarin, F., 2001. Unemployment Duration: An Analysis of Incomplete, Completed, and Multiple Spells in Emilia-Romagna. *Quality & Quantity* 35: 203–230.

Lancaster, T. 1979. Econometric Methods for the Duration of Unemployment. *Econometrica* 47, 939–956.

Lancaster, T., 1990. *The Econometric Analysis of Transition Data*. Cambridge University Press, New York.

Lee, E.T., 1993. *Statistical Methods for Survival Data Analysis*, 2nd edition. Wiley & Sons, New York.

Lippman, S.A., McCall, J.J., 1976, The Economics of Job Search: A Survey, *Economic Inquiry* XIV, 155-189.

Lynch, L., 1989. The Youth Labour Market in the Eighties: Determinants of Re-Employment Probabilities for Young Men and Women. *The Review of Economics and Statistics* 71 (1), 37–45.

Narendranathan, W., Nickell, S., Stern, J., 1985. Unemployment Benefits Revisited. The Economic Journal 95: 307–329.

Narendranathan, W., Stewart, M.B., 1993. Modelling the Probability of Leaving Unemployment: Competing Risks Models with Flexible Base-line Hazards. *Journal of the Royal Statistical Society*, Series C 42, 63–83.

Nickell, S., 1979a. Estimating the Probability of Leaving Unemployment. *Econometrica* 47, 1249–1266.

Nickell, S., 1979b. The Effect of Unemployment and Related Benefits on the Duration of Unemployment. *Economic Journal* 89, 34–49.

Røed, K., Nordberg, M., 2003. Temporary layoffs and the duration of unemployment. *Labour Economics* 10, 381–398.

Salant, S.W., 1977. Search Theory and Duration Data: a Theory of Sorts. *Quarterly Journal of Economic* 91, 39–57.

Sikkel, D., 1985. Models for Memory Effects. *Journal of the American Statistical Association* 80, 835–841.

Silcock, H., 1954. The Phenomenon of Labour Turnover. *Journal of the Royal Statistical Society*, Series A CXVII, 429–440.

SPSS 1997. SPSS® *Advanced Statistics 7.5*, SPSS, Chicago.

Storn, R., Price, K., 1995. Differential Evolution – a Simple and Efficient Adaptive Scheme for Global Optimization over Continuous Spaces, Technical Report TR-95-012, ICSI.

Sudman, S., Bradburn, N. M., 1973. Effects of Time and Memory Factors on Response in Surveys. *Journal of the American Statistical Association* 68, 805–815.

Torelli, N., Trivellato, U., 1988. Modelling Job-Search Duration from the Italian Labour Force Data. LABOUR 2(1), 117–134.

Torelli, N., Trivellato, U., 1993. Modelling Inaccuracies in Job-Search Duration Data. *Journal of Econometrics* 59, 187–211.