Unemployment duration and competing risks: A regional investigation

Chiara Mussida[*]
chiara.mussida@unicatt.it

September, 2007

Abstract

This paper presents empirical findings by applying a duration analysis and competing risks models (CRM) to Lombardy's labour force survey for 2004 and 2005. CR and duration models are becoming increasingly pervasive in applied research to explain the factors determining both the time in a state (i.e. unemployment) and the exit route from the state (i.e. leaving unemployment for a job or non-participation).Starting from a description of the main features and findings on unemployment duration for European countries we add empirical evidence for the Italian context.
For a sample of 823 individuals experiencing unemployment we obtain useful insights on the baseline hazard by applying nonparametric estimation of the survivor function and we estimate four different parametric models within which the Weibull one provides the best fit to the duration data. We find evidence of negative duration dependence and higher hazards of leaving unemployment for married, younger and better educated males. Then, by considering four different competing causes, we estimate CR models to investigate what the characteristics are which drive transitions out of unemployment. Males again present higher employment probabilities, while a consistent proportion of females leave the labour force after a joblessness experience, signalling the presence of a discouragement effect. These findings renew the importance of the kind of analysis conducted, mainly from a policy perspective.

JEL Classifications: D21, O30
Keywords: Unemployment duration, Competing Risks

1 Introduction

How long do individuals spend unemployed? How does the duration of unemployment vary? What are the destination states of the unemployed? Answers to questions such as these are needed for several reasons[1]. Duration and CR models are becoming increasingly pervasive in applied econometrics. By investigating issues like the time in a state (i.e. unemployment) and the exit

[*] The author acknowledges Professor Maurizio Baussola and Professor Colin Cameron for their contribution in the production of this paper, and Professor Richard Dickens for his useful suggestions. The financial support of the Catholic University is gratefully acknowledged. The outcomes and interpretation expressed in this paper are exclusively of the author.
[1] A detailed explanation is in Kiefer (1988).

route from this particular state, they promise a deeper empirical insight into the processes accounting for individual differences in such outcomes. Furthermore, as suggested by literature[2], economists have often found it useful to look at the average length of an unemployment spell in evaluating labour market conditions and in considering the labour market experience of the unemployed. Statistical analyses of unemployment duration are primarily concerned with the probability of leaving unemployment of various durations. The determinants of this escape probability constitute a primary focus of interest and have been studied empirically by, inter alia, Clark and Summers (1979), Lancaster (1979), and Nickell (1979). Furthermore the probability of an unemployed person's finding a job after a certain length of time out of work, and the variation in this probability, are currently of great interest to economists, especially useful from a policy perspective[3], and for the general public alike. While applications of this type are used to inform the policy makers of many foreign countries, in Italy we lack this kind of analysis. For this reason, the aim of this paper is to offer a starting point of analysis by presenting empirical evidence on unemployment duration and CR. The structure of the paper is the following. The second section explains the European context and the main empirical evidence on unemployment. In section 3, we introduce the data description and the main features of Italian unemployment duration. We also provide an application of the duration analysis to Lombardy's labour market, emphasizing the relevance of the methodology applied. The main implications of this analysis are explained in depth. In the last section we introduce the conceptual framework of the CR approach and we apply this technique to the same sample used for the duration analysis. A detailed explanation of the results, and an attempt to advance the main policy implications are given in this last section. Conclusions are provided in section 5. There is little published academic research on the application of the techniques employed. This is one of the reasons that motivate the author to add further

[2]In their work, Baker and Trivedi (1985) offer a detailed description of the literature on this topic.
[3] These measures are often used as parameters to judge the effectiveness of new regulations or, even better, to introduce them. For example, as mentioned by Thomas (1996), in the UK they have been useful to know the observable characteristics of those who are more likely to leave unemployment via part-time jobs.

research to analyse the entire Italian context since these methodologies could provide further useful insights.

2 Unemployment Duration: dynamics and empirical evidence

One of the distinctive features of many European industrialised labour markets is the high incidence of long-term unemployment (LTU). For this reason, as suggested by the OECD, efforts to reduce the length of unemployment spells should be a key element in strategies aimed to reduce overall unemployment[4]. However it seems that, after a sharp increase in unemployment duration, this tendency has lessened, at least in the last few years. Given the relevance of this phenomenon, voluminous empirical evidence has been produced, mainly to try to analyse its causes and consequences. In this section the literature on this topic is reviewed and, where it is possible, additional empirical evidence is provided.

The analysis of LTU has been focused not only on the usual unemployment indicators (typically the unemployment rates), but also on specific indicators summarizing duration distribution. One of these measures is the incidence of LTU on the total unemployment pool (or on the labour force). There are different definitions for LTU, which can be divided into two groups: the statistical and the administrative one. This paper, as it will be clarified below, will adopt the statistical definition of LTU, because it is the survey-based measure of unemployment. More precisely, the Labour Force Surveys of most countries (as well as Italy) ask questions that are designed to find out how long the unemployed have been in that state. Considerable effort has gone into providing a consistent approach to labelling the current labour market state of individuals as unemployed. From the statistical point of view, the International Labour Office (ILO) classifies as LTU individuals unemployed for twelve months or more[5] . Typically those who are currently looking for work are asked how long they have been searching for work. There is obviously no way to check the validity of the answer to this question and given that, as is known, individuals' recall of the length of spells has considerable measurement error as short spells

[4] OECD (2002a) provides a detailed description of the different definitions of long term unemployment and of the feasible policies to reduce its incidence.

[5] This definition is used to produce the OECD unemployment rates, which obviously include the Italian ones.

are often forgotten and there is considerable rounding of answers, it is obvious to expect to have considerable measurement error. Karr (1977) stressed a potential overestimation of LTU incidence due to the fact that the published statistics are based on ongoing unemployment spells; therefore these durations are censored and obviously shorter than the complete ones. On the other hand, there is the length bias problem (Salant, 1977) that, as will be explained below, often leads to an overrepresentation of longer unemployment spells.

What the existing literature suggests is the predominance of the length bias effect over the censoring problem; therefore, the effective incidence of LTU is lower than that published.

Empirical evidence (mainly Machin and Manning, 1999) helps to better understand the causes and consequences of LTU. The importance of this phenomenon is underlined by statistics on its incidence in European countries[6]. Great interest lies also in the relationship between the overall unemployment rate and the incidence of LTU for the OECD countries analysed. As emphasized by literature, there is a positive relationship between the two variables, but care must be taken in interpreting this result. However, the most important fact is the evidence of a delay in the incidence of LTU with respect to changes in the employment cycle[7].

The availability of LTU incidence allows a detailed analysis of the labour market, mainly because at a given unemployment rate it can lead to two conflicting situations. On the one hand the overall population could experience brief unemployment spells. On the other hand a group of “hard-core unemployed” could bear the total weight of the unemployment. For sure these opposite situations imply different policies, both in terms of active interventions (to increase the employment opportunities for the unemployed), but also in terms of economic and social interventions to assist the people in need[8].

[6] In 2000 the average incidence of the LTU in the industrialised countries has been by 30%, but the features and the dynamics of this phenomenon are heterogeneously distributed over the countries. Italy and Germany are the countries with the highest incidence (greater than 50%), while United States and Canada are the states with the lowest incidence.

[7] This point is theoretically explained in Machin and Manning (1999).

[8] In their work, Payne and Payne (2000), and Paggiaro (2001) provide specific approaches to identify the people more exposed to the risk of unemployment.

As regards the causes of unemployment in the OECD countries, Machin and Manning (1999) stress the role of the changes in the inflows and outflows from unemployment, but also of the duration dependence. They mainly conclude that the increases in the incidence of LTU and high unemployment rates have had a common cause: the collapse of exit rates from unemployment at all durations. Their finding is consistent with the fact that duration dependence and inflow rates do not seem to have changed very much over time. However, as will be shown in the next paragraph, the results for the Italian labour market analysis lead to different interpretations of the causes of LTU.

2.1 Inflows and transitions from the unemployment state

Analysis of the unemployment stock in terms of its dynamic components – mainly inflow and outflow rates - is essential to evaluate if the findings of Machin and Manning (1999) are confirmed for the Italian labour market[9]. For this purpose it is useful to introduce the results obtained by the previous Italian labour force survey. First of all, the incidence of LTU is related to the time evolution of the number of unemployed of short and long duration. Analysis of the unemployment pool's duration can provide evidence on the inflow dynamics in this state. By dividing the unemployment into short (inflow over the last year) and long term (twelve months or more) durations it is possible to study the relationship between these two components. This kind of analysis covered the time period 1993-2002. One of the most interesting patterns is the inflow one, which showed a sharp decrease, reducing its value by approximately half. Furthermore, in every geographical division examined, a decreasing trend was noticed. This confirms the inverse relationship between the inflow rate and the incidence of unemployment, at least for the short run[10]. This result obviously

[9] They conclude, as mentioned above, that the raise in the LTU incidence is mainly due to a reduction of the exit rates from unemployment, given that duration dependence and inflow rates do not seem to have changed very much over time.
[10] For further details on the analysis described, see Contini and Trivellato (2005).

contrasts with the findings of Machin and Manning (1999)[11]. As regards the medium term effects, the analysis is focused on the LTU series which, as will be explained later in the paper, lags behind that of short duration. Given that the LTU are a fraction of previous years' inflow that didn't change their state, the inflow reduction is also going to affect this category, but with a time lag.

Looking at the destinations' states, there are three possible alternatives: to become employed, remain unemployed, or leave the labour force. The evidence is towards a substantial stability of these historical series. This is true for all the geographical partitions and for both the LTU and the short term unemployed. Also this result contrasts with that of Machin and Manning, because they inferred a reduction in the outflow rates in response to a LTU increase.

It is quite interesting to extend the analysis of the unemployment duration for the time span covered by the actual survey. It is important to underline that, as will be explained later in the paper, it is based on quarterly data for the years 2004 and 2005 and related to Lombardy's labour market.

3 Unemployment Duration Data

The aim of this paragraph is to analyse unemployment duration for one of the most important Italian regions, Lombardy. Analysis is limited to one region because there was no time to collect national data. I have already requested this data and as soon as I receive all the statistics I will extend the methodologies applied in this paper to the Italian data. However, acknowledging the great potential of this analysis and given that one-sixth of the population of Italy lives in Lombardy (9.475.202 people) and it is one of the three richest regions in Europe, with a per capita gross domestic product that is 30 percent higher than the rest of Italy (and represents 20 percent of Italian Gross

[11] This statement comes from the fact that Machin and Manning (1999) – as emphasized in note 9 – suggest that the raise in the LTU is mainly due to a reduction of the exit rates from unemployment, given that duration dependence and inflows rates do not seem to have changed very much over time, while the Italian case (data covering the time span 1993-2002) suggests a decreasing inflow trend causing the raise of LTU incidence (inverse relationship between inflow rates and unemployment incidence).

Domestic Product[12]) - Lombardy is one of the most economically advanced regions in Italy and the European Union, I think that the results obtained in this application are significant and representative for the future extension mentioned.

Interest lies mainly in the pattern of survival times. Specifically, after checking for observable individual characteristics, does success feed on itself in the sense that the conditional probability of leaving unemployment is lower for individuals who experienced longer spells?

These important questions are answered by using regression models with dependent variable defined to be the duration of time (months) until the individual leaves the state of unemployment. In the estimated model this variable is called spell and, as every dependent variable used in this kind of estimation, cannot be negative and from the stochastic process theory it is likely to be distributed as exponential (or a generalization of exponential). Figure 1 illustrates unemployment duration distribution across individuals who experienced unemployment in the two years analysed (2004 and 2005) and left this state as suggested by the survey. To better get the details the graph only includes the unemployment experiences lasting up to 100 months, since only a few cases reported higher durations. Anyway, the analysis that follows refers to all the sample of unemployed[13]. The mean spell lasts 20 months, while the shortest duration is one month (or less)[14] and the longest 417 months. As can be seen from the histogram, 1 month of unemployment is the event with the highest density.

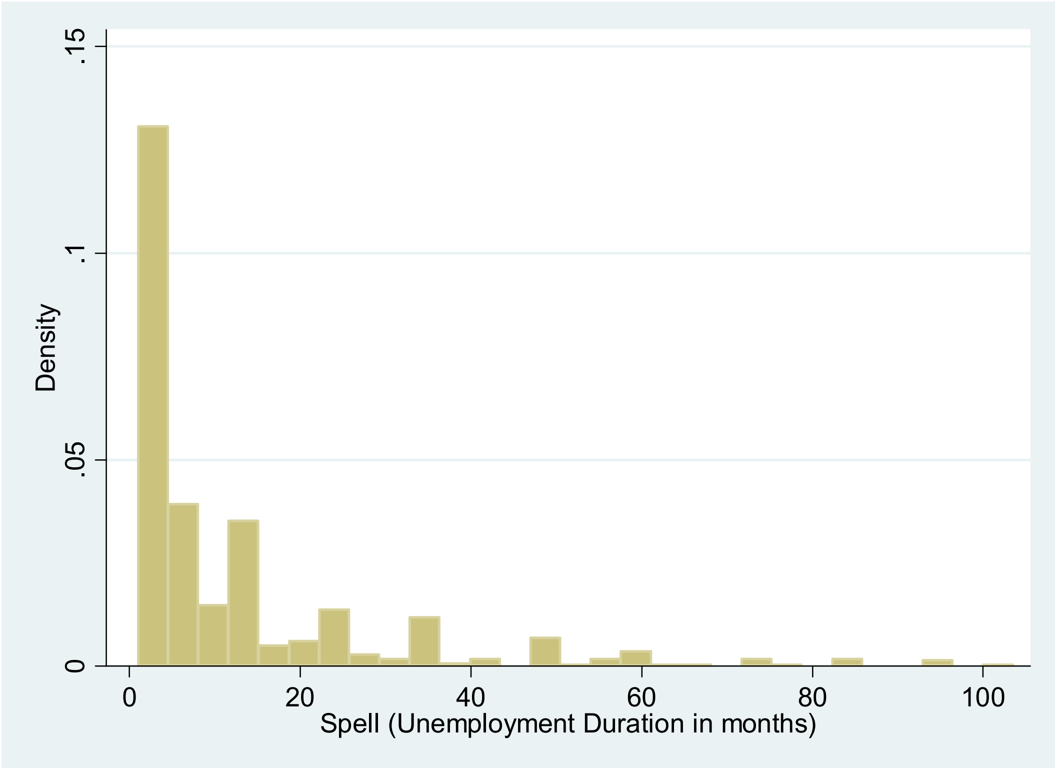
---

[12] Data source: Istat, labour force survey for 2004 and 2005, Rome.

[13] The exclusion of the individual experiencing higher durations – as it will be shown in table 2 (Kaplan-Meter Survivor Function estimates) - would imply the lost of 37 individuals (4.5 % of the total sample). For this reason, to maintain the precision of the estimates, and to offer an overall picture of the Lombardy's unemployment, all the unemployed enters the models' estimations (overall sample size of 823 individuals).

[14] In order to increase the estimates' precision and to avoid the exclusion of individuals who experienced unemployment for less than a month, they have been included in the sub sample analysed. It is noteworthy to underline that this inclusion does not bias the basic estimation results.

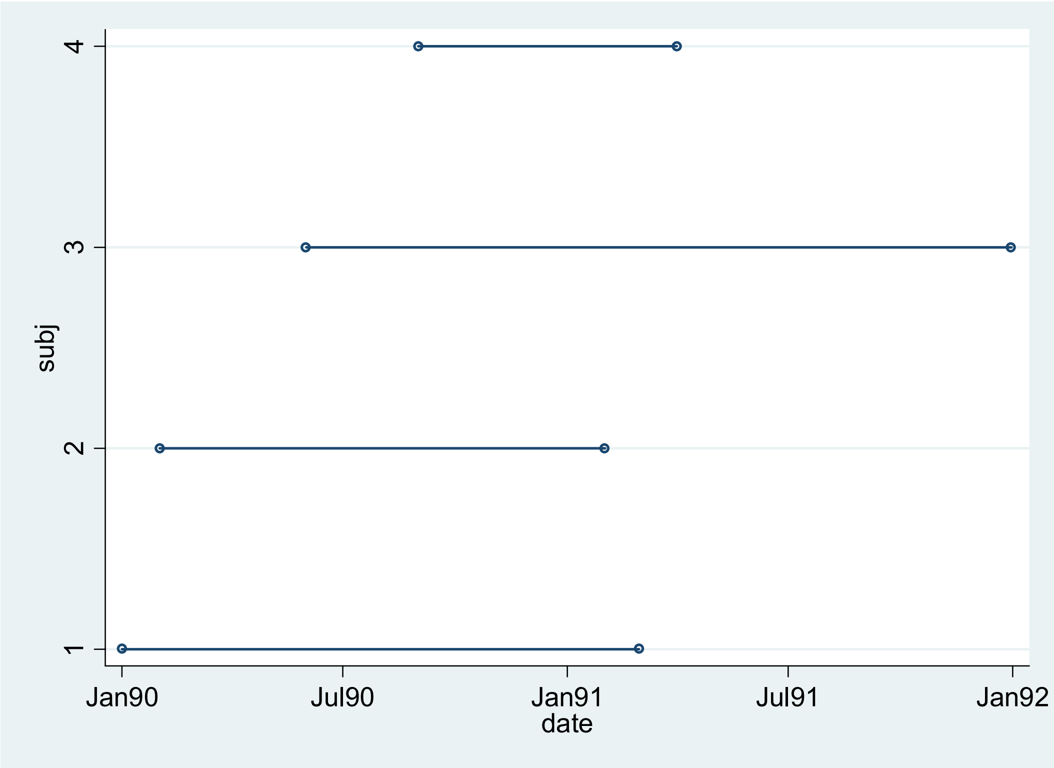
Figure 1: Histogram of unemployment spell distribution



Source: Elaborations from Stata

The analysis is also complicated because the data are censored. For a sub sample of people interviewed it is not possible to build up a complete unemployment story since they are still unemployed at the time of the survey.

Figure 2: Duration data



Source: Elaborations from Stata

From the graph above it can be inferred that the only individual with an uncompleted unemployment spell at the time of the survey is subject 3 (right-censored spell), while for the others the duration can be recorded (uncensored spells).

More precisely, defining duration requires a time origin (a beginning), a time scale, and a precise definition of the event ending the duration. In the sample analysed, different individuals will have different time origins for the durations they experience. The unemployment spells could begin at any date; the beginning date is the time origin for the spell. The duration of a spell is its length[15]. The length of time people spend on average looking for work is also an important index of economic welfare. It would seem that such information could be readily obtained by consulting the labour force survey compiled by the Italian National Institute of Statistics (Istat). However, like many other labour market

15 A detailed explanation is contained in Kiefer, N.M (1988), *Economic Duration Data and Hazard Functions*, Journal of Economic Literature, Vol. 26, no. 2, pp. 646-679.

surveys, sometimes it provides little direct information about the length of unemployment spells. To better understand this statement it is important to introduce a fundamental theoretical consideration related to this type of analysis. Looking again at the above graph, each horizontal line represents a spell of unemployment. The length of each line indicates the length of a completed spell ($\hat{S}$), the random variable that interests search theorists and the public[16]. Four realizations of this random variable are shown in the above graph. The National Statistical Institute continuously surveys the unemployed and the survey structure allows each individual's spell length to be built up, measuring in this way also the length of spells in progress up to the date of the last survey. In the graph, only one spell was in progress when the survey was conducted (in the example this date is January 1992). What the Istat observes of this spell is only part of its length. This partial length is called the length of an interrupted spell ($\hat{T}$).

The theoretical explanation which follows assumes for simplicity that the economic conditions are stable. That is, a spell is likely to start at any time, and its length is assumed to be drawn from the same distribution regardless of when the spell begins. As the above graph suggests, the full length of any spell captured by the survey will exceed the partial length measured by the Istat ($S_i \geq T_i$). Since under stable economic conditions the intersection of a spell with the survey is equally likely to occur at any point on the length of the spell, captured spells are on average halfway through their full length at the time of the survey. This phenomenon is defined as "interruption bias". On the other hand, as suggested by the theory and by the above graph, it is spells with longer than average full length that are more likely to be in progress at the time of the survey. This is the "length-biased population" phenomena. Because the effects of interruption bias and length-bias conflict, no statement can be made in general about which of the two random variables (complete or interrupted spell) has the larger mean. However Salant (1977) provides a general relationship between the means.

[16] Salant (1977) provides a complete explanation in his work on the search theory and duration data.

A feature of the completed spell-length density that determines which of the two effects predominates is its failure or escape rate, indicating the hazard that a spell will terminate in a given week given that it has not ended before then.
As a consequence, three useful propositions from “renewal theory” follow:

- ✓ if the probability of escape rises with time unemployed, E(T) < E(S);
- ✓ if the probability of exiting is constant, E(T) = E(S). The effects of length-bias and interruption-bias exactly offset each other and the spell-length variates are exponentially distributed;
- ✓ if the probability of escape falls with time spent in the unemployment state, E(T) > E(S).

These propositions will be needed later in the paper when the model estimated for the duration of the Lombardy’s unemployment will be discussed. To conclude this theoretical explanation, one point has to be made. An examination of the data collected by the labour market survey reveals that the odds of escape decline with time unemployed. As inferred from the first graph, a large fraction of the one-month unemployed disappear (to employment or out of the labour force) before they can be classified two-months unemployed.

3.1 Italian unemployment duration: evidence from micro data and the actual Survey

Before introducing the survey employed in this paper and the features of the sample analysed, it is necessary to describe the advantages of the Italian labour market survey and the main results obtained for unemployment duration with the micro data for the time period 1993/94 – 2002/03. This time span is covered by the “quarterly labour force survey” that, as will be explained below, has been modified since 2004. The availability of historical series of the unemployment rates and of the LTU incidence is essential to identify the presence of dynamic relationships between these two series.

First of all, an increase in LTU incidence has been inferred. This phenomenon rises at high rates, but without showing any regularity. More precisely, this increase continued until the 90s, and then it started to slow down. Since 2000 it has been characterised by a trend reversion towards lower incidences.

Looking at the relationship between LTU and the unemployment rate, the most remarkable feature is that long-term unemployment displays anti-clockwise loops or, alternatively, it lags behind actual unemployment. More precisely, starting from the peak of the cycle as unemployment rises the share of LTU actually falls at first but then rises. Once the trough is reached and unemployment starts to fall the proportion of LTU continues to rise for a while but then falls. The consequence of this is that for a given level of unemployment, the incidence of LTU is generally higher in the recovery than the slump. This behaviour is consistent with an unemployment rate mainly linked to the inflows, given that the outflow rates do not seem to have changed much over time[17].

The Italian labour force survey is the main source of statistical documentation on the labour market. For this reason it is mainly focused on measurement of employment and unemployment, but it has also to provide ways and degrees of labour market participation. All these requirements have led to a new structure which currently characterises the labour force survey.

Like the previous survey's structure (called quarterly labour force survey), the main objective of the new one is to produce official estimates of the number of people employed and unemployed (or in search of a job). For this reason it is necessary to divide the population (of working age which includes people aged 15 or over) into three groups in order to cover all the possible alternatives: employed, unemployed and out of the labour force. This classification is based both on criteria provided by the International Labour Office (ILO) and on suggestions of the EU regulations.

Under these definitions, an individual is classified as being employed if he has worked at least an hour in the week of reference, which in general is the week which precedes the survey. The definition of unemployed is based on detailed

[17] This hypothesis is rejected by the work of Machin and Manning (1999), but it characterizes the Italian results.

requirements, applied to people aged 15 or over and under 74[18]. More precisely (i) the individual does not have a job; (ii) s/he is available to start working within the two weeks following the interview; and (iii) s/he has looked for a job (using one of the job search actions listed in the questionnaire) in the past four weeks. Finally, a person is considered as "out of the labour force" if he didn't work and he did not look for work (again in the last four weeks)[19].

It is important to emphasize that, as mentioned above, the survey has been modified many times, but the main changes were introduced from 2004. One of these changes was the introduction of an upper age limit to the estimation of the people in search of a job. This limit was fixed at 74 years. However, the main differences in the labour market participation estimates will be given by the methodological innovation introduced by the new survey (defined as the continuous labour force survey) which is the continuity of the interviews - conducted every week of a quarter - instead of a single survey, concentrated in a reference week in the quarter. Additional information will also come from the definition of the unit of survey. This is defined as de facto family – cohabiting persons, tied by wedding, relationship, affinity, adoption, protection or affective ties. The criteria employed to identify a de facto family is cohabitation and the relationship or affective ties which tie its members. The people selected by the survey are all the family members resident in Italy, excluding permanent members of military and religious institutes. The new survey allows a wider range of relationship ties which tie the family members. This is the most relevant innovation introduced by the continuous survey in terms of unit.

In addition, it surveys also the wedding's year. With this information it is possible to build up the ties within the reference individual and the other family members, increasing in this way the statistical base available to study the relation between families and labour market.

---

[18] For a detailed explanation of these criteria see the 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.

[19] It is note worthy to underline that the same definitions are applied, as explained in Akerlof and Main (1980), by the U.S. Bureau of Labor Statistics.

The data sets used in this paper were collected by the new survey described. More precisely, it covers eight quarters for the years 2004 and 2005 and it is a rotated panel[20].

However, before proceeding with the sub sample description, it is essential to provide additional information about the sample survey.

The new survey was introduced, as mentioned above, mainly to satisfy European Union criteria. Two main changes: the first is the survey regularity (it is continuously distributed over the quarter) and the second one is related to the precision of estimates. It is necessary to underline the following points:

- ✓ The survey is continuous and it provides both quarterly and annual results;
- ✓ The information collected is related to the reference week (which precedes the survey)
- ✓ the quarters and years of reference are composed of 13 and 52 weeks respectively.

A detailed explanation of the survey scheme is needed to really understand the features of the samples surveyed. The quarterly samples interviewed partially overlap, following a specific rotation scheme. More precisely, a family was interviewed for two consecutive surveys and, after leaving the sample for two quarters, was interviewed for another two consecutive quarters. This is defined as a (2-2-2) rotation scheme. As shown in the table below, it implies a 50% overlapping of the theoretical sample to a quarter of distance, a 25% overlapping to three quarters, a 50% to four quarters, and a 25% to five quarters.

Table 1: The rotation scheme (2-2-2)

Quarter	**Rotation Groups**						
I quarter year a	A4	B3			E2	F1	
II quarter year a		B4	C3			F2	G1

[20] The reasons why it is a rotated panel are well explained in this paragraph. However, the great potential of this survey is the possibility to obtain both quarterly and yearly labour market indicators.

III quarter year a	C4	D3			G2	H1			
IV quarter year a		D4	E3			H2	I1		
I quarter year a+1			E4	F3			I2	J1	
II quarter year a+1				F4	G3			J2	K1

Source: Istat, Italian National Institute of Statistics.

The quarterly sampling design is composed of two stages, with a stratification of the units of the first stage; the first stage units are municipalities, while the second stage ones are families. In every Italian province municipalities are classified in two different categories: a municipality with a population which exceeds a fixed threshold, defined as auto-representative municipalities (Ar); while the remaining are defined as non-auto-representative (Nar). Every Ar municipality is included in the sample interviewed, while a stratification based on the population is applied for the Nar municipalities' selection. A sample of families is then selected from the identifying list provided by every drawn municipality. Every family member is interviewed. The main difference between the two stages mentioned is that while for families a rotation scheme is applied (table 1), drawn municipalities do not change over the time.

To better understand the great usefulness of the survey described it is important to clarify another issue. To obtain a representative sample of the total population it is necessary to take into consideration the sampling weights used by the labour force survey. These weights denote the inverse of the probability that the observation is included due to the sampling design and a correction is applied to guarantee equality between the total actual population (classified by sex and age groups) and the relative sample estimates. In this way representative national estimates can be obtained. This is obviously a great advantage because it allows detailed labour market analysis.

More precisely, in every quarter the national survey interviews 1430 municipalities, including almost 70.000 families for a total of approximately 176.000 individuals[21].

[21] A detailed description of the survey is included in Istat (Istituto Nazionale di Statistica, 2006), *La rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione*, Metodi e Norme n. 32, Roma.

The data set used in this paper, sourced from Istat[22], is related to Lombardy, which is a Northern Italian region, characterised by a high level of GDP and population. The continuous labour force survey, as mentioned above, covers 2004 and 2005 (eight quarters data). More precisely, in order to build up and examine the individuals' unemployment history a sub sample of the regional survey was selected. It is composed of the people who experienced unemployment. A proper description of this sub sample composition will be given later in the paper. By exploiting the detailed information provided by the new survey it was possible to compute the unemployment rates conditional on individual characteristics, such as sex, age groups, and education. These indicators, shown in table A1 (appendix) provide detailed insights on this phenomenon, and allow a deeper analysis aimed at discovering the characteristics of the unemployed. This information is undoubtedly useful for policy analysis. This also enables the computation of detailed transitions probabilities between the labour market states[23].

3.2 An application to Lombardy's labour market: Nonparametric and Parametric Duration Data Models

The use of duration (or survival) models is relatively recent in economics although they have been extensively used in engineering and biomedical research for many years. Survival analysis, the term used for this approach within the biomedical tradition, is concerned with a group of individuals for whom a point event of some kind is defined. This event is often defined as a failure. The event of failure occurs after a length of time called the failure time and can occur at most once for any individual or phenomenon under inspection. One example of these failures is the duration of a period of unemployment experienced by an individual.

[22] The data set used is the labour force survey for 2004 and 2005 collected by Istat (Italian National Institute of Statistics), Rome.

[23] The availability of detailed information also for the Italian labour market will allow a detailed analysis of the causes and characteristics of the national unemployment. Furthermore, the transition probabilities' might provide additional insights on the labour market mobility.

To really understand the technique used and the models estimated in this paper it is fundamental to introduce the basic concepts related to this kind of analysis[24].

Econometricians use the term spell length to describe the time occupancy or duration of a given state. The spell length is usually represented by a random variable, which is denoted by T. T is assumed to be a continuous random variable and we assume a large population of people enter some given state at a time identified by T=0. The calendar time of state entry need not be the same for all individuals. T is thus the duration of stay in the state. The population is assumed to be homogeneous implying that everyone's duration will be a realisation of a random variable from the same probability distribution.

If we define the probability that a person who has occupied a state for a time t leaves it in the short interval of length Δt as:

$$prob[t \leq T \leq t + \Delta t \,|\, T \geq t] \tag{1}$$

the conditioning event ($T \geq t$) in (1) is the event that the state is still occupied at t.

In other words, the conditioning event is that the individual has not left the state before time t. If we divide (1) by Δt we obtain the average probability of leaving per unit of time period over a short interval after t. By taking this average over shorter intervals we can define:

$$\theta(t) = \lim_{\Delta t \to 0} \frac{prob[t \leq T < t + \Delta t \,|\, T \geq t]}{\Delta t} \tag{2}$$

as the hazard function. It is the instantaneous rate of leaving per unit of time period at *t*. The interpretation of $\theta(t)\Delta t$ is the probability of exit from a given state in the short interval of time Δt after *t*, conditional on the state being occupied at time *t*.

---

[24] For more details on Duration Data Models, see Cameron and Trivedi (2005, chapter 17).

Define $prob[T < t] = F(t)$ and note that $prob[T \geq t] = 1 - F(t)$, also identified as *S(t)*. One minus the distribution function is an expression quite common in applications involving duration or survival data. It is defined as survivor function since it gives the probability of survival to time *t*. Of any inflow group (cohort) of persons entering unemployment, it gives the proportion of the population who stay at least *t* years.

Given that the derivative of the cumulative distribution function is nothing more than the density function:

$$\frac{\Delta F(t)}{\Delta t} = f(t) \tag{3}$$

We can use these results to re-write the hazard function as:

$$\theta(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)} \tag{4}$$

this is, as mentioned above, the instantaneous probability of leaving a state conditional on survival to time *t*.

Prior to introducing the models for duration data, it is worth noting that $\theta(t)$ can be expressed in an alternative way. More precisely, the log of the survivor function can be written as:

$$\log[1 - F(t)] \tag{5}$$

And the derivative of this expression with respect to time, by coding $[1 - F(t)]$ as z and by using the chain rule can be written as:

$$\frac{d\log[1 - F(t)]}{dt} = \frac{d\log[z]}{dz}\frac{dz}{dt} = \frac{1}{z}[-f(t)] = \frac{-f(t)}{1 - F(t)} = -\theta(t)\,. \tag{6}$$

This can be expressed as

$$\theta(t) = \frac{-d\log[1 - F(t)]}{dt} = \frac{-d\log[S(t)]}{dt} \tag{7}$$

or

$$\frac{d\log[1-F(t)]}{dt} = -\theta(t) \tag{8}$$

A final related function, which is important in this literature, is the cumulative hazard function or integrated hazard function, defined as:

$$\wedge(t) = \int_0^t \theta(s)ds = -\log(1-F(T)) = -\log(S(t)) \tag{9}$$

For any choice of distribution of T, it can be shown that the transformation $\wedge(t)$ is unit exponentially distributed and log $\wedge(t)$ is extreme-value distributed.

Before introducing regressors it is convenient to present an estimate of the distribution of unemployment duration. This procedure is the nonparametric estimation of the survivor function and it is useful for descriptive purposes, but also to obtain insights on the baseline hazard. The standard procedure in duration analysis is to estimate the survivor function, $S(t) = 1 - F(t)$, controlling for censoring by using the non parametric Kaplan-Meyer estimate

$$\hat{S}(t) = \prod_{j|t_i \leq t} \left(\frac{n_j - d_j}{n_j} \right) \tag{10}$$

where n_j is the size of the risk set at time t_j and d_j is the number of observation spells completed at time t_j. The computation of this measure, along with the 95% confidence bands, is contained in the following table

Table 2: Unemployment Duration: Kaplan-Meyer Survivor Function Estimates

Time	Beg. Total	Fail	Net Lost	Survivor Function	Std. Error	[95% Conf. Int]	
1	823	41	139	0.9502	0.0076	0.9329	0.9631
2	643	16	63	0.9265	0.0094	0.9057	0.9429
3	564	14	55	0.9035	0.0110	0.8795	0.9230
4	495	18	17	0.8707	0.0131	0.8426	0.8940
5	460	20	11	0.8328	0.0150	0.8011	0.8600
6	429	25	12	0.7843	0.0170	0.7488	0.8154
7	392	5	11	0.7743	0.0173	0.7381	0.8061
8	376	15	10	0.7434	0.0184	0.7053	0.7774
9	351	8	6	0.7265	0.0189	0.7864	0.7615
10	337	8	6	0.7092	0.0194	0.6692	0.7453

11	323	9	4	0.6895	0.0200	0.6484	0.7267
12	310	43	34	0.5838	0.0219	0.5495	0.6353
13	233	5	2	0.5811	0.0221	0.5364	0.6231
14	226	3	1	0.5734	0.0223	0.5284	0.6157
15	222	4	6	0.5630	0.0225	0.5177	0.6058
16	212	2	2	0.5577	0.0226	0.5123	0.6007
17	208	0	1	0.5577	0.0226	0.5123	0.6007
18	207	7	2	0.5389	0.0229	0.4929	0.5826
19	198	1	0	0.5362	0.0230	0.4910	0.5799
20	197	2	7	0.5307	0.0230	0.4845	0.5747
21	188	0	4	0.5307	0.0230	0.4845	0.5747
22	184	2	1	0.5249	0.0232	0.4785	0.5692
23	181	0	1	0.5249	0.0232	0.4785	0.5692
24	180	20	14	0.4666	0.0240	0.4190	0.5128
25	146	0	3	0.4666	0.0240	0.4190	0.5128
26	143	0	1	0.4666	0.0240	0.4190	0.5128
27	142	1	1	0.4633	0.0240	0.4156	0.5096
28	140	0	5	0.4633	0.0240	0.4156	0.5096
30	135	3	1	0.4530	0.0242	0.4050	0.4997
32	131	1	0	0.4496	0.0243	0.4014	0.4964
33	130	0	1	0.4496	0.0243	0.4014	0.4964
35	129	0	2	0.4496	0.0243	0.4014	0.4964
36	127	11	19	0.4106	0.0249	0.3617	0.4589
39	97	0	2	0.4106	0.0249	0.3617	0.4589
40	95	1	2	0.4063	0.0250	0.3572	0.4548
41	92	0	1	0.4063	0.0250	0.3572	0.4548
42	91	0	1	0.4063	0.0250	0.3572	0.4548
47	90	0	2	0.4063	0.0250	0.3572	0.4548
48	88	7	9	0.3740	0.0258	0.3235	0.4243
50	72	0	1	0.3740	0.0258	0.3235	0.4243
51	71	1	0	0.3687	0.0260	0.3180	0.4194
57	70	2	3	0.3582	0.0263	0.3070	0.4096
60	65	2	8	0.3472	0.0266	0.2954	0.3993
63	55	0	1	0.3472	0.0266	0.2954	0.3993
68	54	0	1	0.3472	0.0266	0.2954	0.3993
72	53	1	3	0.3406	0.0269	0.2884	0.3934
73	49	0	1	0.3406	0.0269	0.2884	0.3934
78	48	0	1	0.3406	0.0269	0.2884	0.3934
84	47	3	2	0.3189	0.0279	0.2649	0.3740
93	42	0	1	0.3189	0.0279	0.2649	0.3740
95	41	0	2	0.3189	0.0279	0.2649	0.3740
96	39	1	0	0.3107	0.0284	0.2560	0.3668
100	38	1	0	0.3025	0.0288	0.2472	0.3596
108	37	0	3	0.3025	0.0288	0.2472	0.3596
110	34	0	1	0.3025	0.0288	0.2472	0.3596
111	33	0	1	0.3025	0.0288	0.2472	0.3596
120	32	2	5	0.2836	0.0299	0.2265	0.3432
123	25	1	0	0.2723	0.0308	0.2138	0.3339
124	24	0	1	0.2723	0.0308	0.2138	0.3339
129	23	0	1	0.2723	0.0308	0.2138	0.3339
132	22	0	2	0.2723	0.0308	0.2138	0.3339
141	20	0	2	0.2723	0.0308	0.2138	0.3339
153	18	0	1	0.2723	0.0308	0.2138	0.3339
155	17	0	1	0.2723	0.0308	0.2138	0.3339
156	16	0	1	0.2723	0.0308	0.2138	0.3339
189	15	0	1	0.2723	0.0308	0.2138	0.3339
204	14	0	1	0.2723	0.0308	0.2138	0.3339
228	13	0	1	0.2723	0.0308	0.2138	0.3339
240	12	1	0	0.2496	0.0356	0.1830	0.3216
252	11	0	1	0.2496	0.0356	0.1830	0.3216
264	10	1	2	0.2246	0.0399	0.1518	0.3064
268	7	0	1	0.2246	0.0399	0.1518	0.3064
288	6	1	1	0.1872	0.0477	0.1047	0.2882
309	4	0	1	0.1872	0.0477	0.1047	0.2882

336	3	0	1	0.1872	0.0477	0.1047	0.2882
360	2	0	1	0.1872	0.0477	0.1047	0.2882
417	1	1	0	0.0000	.	.	.

Source: Elaborations from Stata

The calculation of the survivor function for the first row of the table is illustrated to better understand the methodology just introduced. The risk set at the start of the time is 823 (individual unemployed). Therefore n_j = 823. In the first month 41 individuals leave unemployment implying d_j = 41. At the start of the first month the survivor function is equal to $\frac{823-41}{823}=0.9502$. This number indicates people staying in unemployment conditional on having survived at the start of the first month (95.02% of the initial risk set). The survivor function declines rapidly up to 24 months of duration time, then the reduction is steady. It is noteworthy to underline that the information on the unemployment durations are available from the questionnaire: there is a question that asks how long (months) the unemployment duration is or has been for every individual. In this way, it has been possible to build up this variable (labelled spell for the models estimates) for every people selected[25]. These computations obviously have allowed more reliable information and a greater estimates' precision.

However, for many economical purposes the analysis of the hazard functions may be more interesting than the survivor one. For this reason the data showed above could be used to compute hazard rates using the Kaplan-Meier hazard formula. This is defined as

$$\theta(T_j)=\frac{d_j}{n_j} \qquad (11)$$

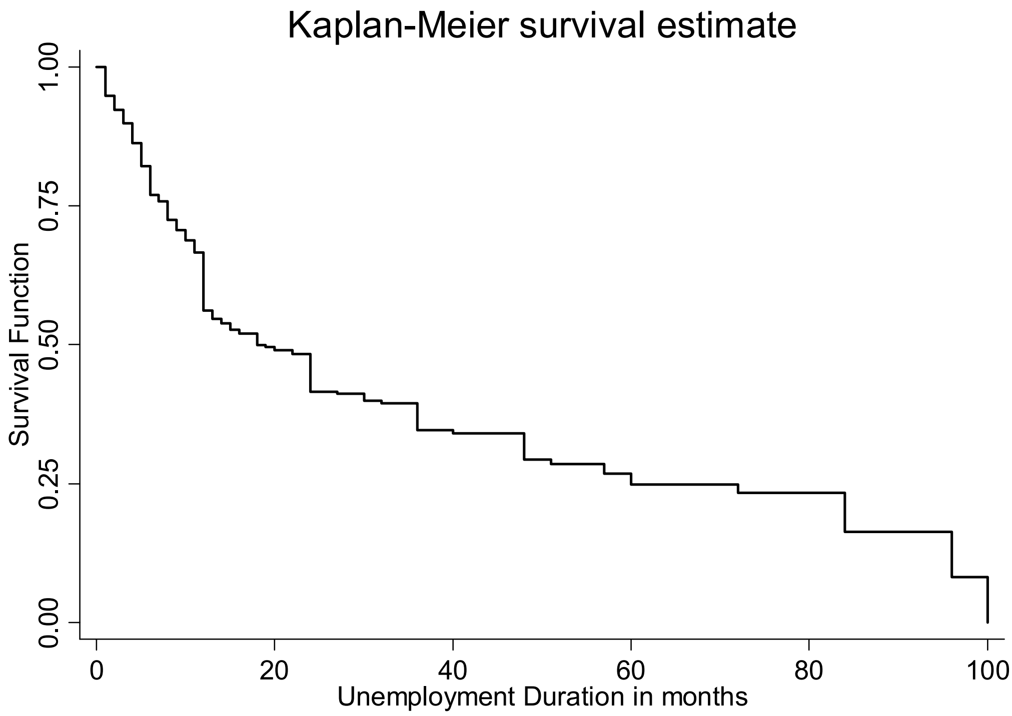
This is a non-parametric empirical approach that imposes no restrictions on the data in the way that the parametric (i.e. Weibull, exponential or Log Logistic) models do. Given the data above, the hazard rates for the first and the second spell could be computed as follows:

$$\theta(T_1)=\frac{41}{823}=0.049\,; \qquad \theta(T_2)=\frac{16}{643}=0,024$$

[25] As already mentioned, every individual with an unemployment experience has been analysed. This is the main criteria for the sample's selection.

The interpretation for the first hazard estimate is that there is a 4.9% chance of exiting the state in the first month. For the second one it can be shown that conditional on surviving to the second month, there is a 2.4% chance of exiting the unemployment. Lastly, useful insights on the baseline hazard can be obtained by plotting the Kaplan-Meyer hazard rates.

Figure 3: Estimated Survival Function with no regressors: nonparametric Kaplan Meyer estimates



Source: Elaborations from Stata

The estimated survivor function (vertical axis) with no regressors, along with the 95% confidence bands, is presented in the above figure. Again, only unemployment experience up to 100 months (horizontal axis) have been included to better get the details since – as it will be explained - there is no much action above this threshold. Anyway, for the same reasons stated before, the estimations have been carried out on the total Lombardy's unemployment sample. For this reason it is noteworthy to underline that the total time period covered by the unemployment duration reaches 417 months. However, the estimates become imprecise after about 100 months (less than 9 years), a consequence of the sample including many recent entrants and relatively few unemployment spells with long durations of either complete and incomplete spells[26].

After checking for the nature of the hazard function it is essential to choose the distribution and the related parametric model which best fits the data. For Lombardy's labour market estimates the proper option is the Weibull distribution. This choice comes from a comparison between different parametric models, within which the Weibull one provides the best fit (primarily in terms of Log likelihood, but also in terms of significance of coefficients and sign). Table A2 in the appendix shows the estimates obtained for all the models considered, which were the exponential, the proportional hazard model (also defined as the Cox PH model)[27] and the log-logistic model.

In contrast to the exponential distribution which is invariant to the duration[28], the Weibull monotonically increases or decreases in duration depending on certain parameters values. Plotting the hazard function, as it will be shown below, can provide useful insights into how the exit probability behaves with state duration. The nature of the relationship between the hazard and the duration is known as duration dependence. A more detailed explanation of this parameter will be given with the estimation results.

The survivor function for the Weibull distribution can be written in this way

---

[26] This is a typical feature of the Italian labour market. In this application we find people experiencing long unemployment durations. It is important to stress that, even if the survey classifies these as joblessness, they probably reflect other disadvantaged labour market's categories. The availability of the entire Italian dataset will allow a more precise definition of these individuals.

[27] By introducing a covariates' function, the hazard function of a PH model can be expressed as: $\theta(x;t) = k_1(x)k_2(t)$, where k_1 and k_2 are the same function for all individuals. The baseline hazard is common to all units in the population and does not vary across individuals. Individual hazards differ proportionately based on the realization of the covariates. The model is defined a PH model because for two individuals with regressor realizations x_1 and x_2, the hazards for two individuals are in the same ratio $k_1(x_1)/k_2(x_2)$ for all t. The proportionate effect of x on the hazard is the same for all dates. This is a relatively restrictive form and it should be noted that there is no obvious reason why hazards should be proportional with duration data from economic applications. This is the main reason which leads to the choice of a less "restrictive" model for the application done in this paper.

[28] The hazard of the exponential distribution is constant and thus independent of time. It is completely described by the parameter θ which is defined as the scale parameter. This definition comes from the fact that each unique value of θ determines a different exponential distribution thus implying the existence of a family of exponential distributions. Given that the distribution has a constant hazard it has been considered quite restrictive and applied economists have looked at alternative distributions.

$$S(t) = \exp\left[-\left(\lambda t\right)^{\alpha} \right] \tag{12}$$

And the cumulative distribution is given by

$$F(t) = 1 - \exp\left[-\left(\lambda t\right)^{\alpha} \right] \tag{13}$$

The derivative of this expression with respect to t (time spent into the unemployment) yields the density function which, in this case, is

$$f(t) = \alpha \lambda^{\alpha} t^{\alpha - 1} \exp\left[-\left(\lambda t\right)^{\alpha} \right] \tag{14}$$

The hazard function can be written as follows

$$\frac{f(t)}{1 - F(t)} = \theta(t) = \alpha \lambda^{\alpha} t^{\alpha - 1} \tag{15}$$

This is the Weibull family of distributions, which became popular also for the simplicity of expressions. In contrast to the exponential case, the hazard is not constant and can either rise or fall with duration. λ and α are known as the scale and the location parameter respectively. The behaviour of the hazard, as mentioned above, depends on the shape parameter α. More precisely, from the hazard function expressed in (15):

- ✓ if $\alpha > 1$ this implies $\partial\theta/\partial t > 0$ (positive derivative of the hazard function with respect to time) and suggests an increasing hazard rate and positive duration dependence;
- ✓ if $\alpha = 1$ this implies $\partial\theta/\partial t = 0$ and suggests a constant hazard rate and no duration dependence (this leads to the exponential case)
- ✓ if $\alpha < 1$ this implies $\partial\theta/\partial t < 0$ and suggests a decreasing hazard rate and negative duration dependence.

One of the reasons why the Weibull is more popular among economists than the exponential distribution is because it allows the analysis of duration data characterized by either positive or negative dependence. However its limitation is that it only allows for constant (leading to the exponential model) and increasing or decreasing hazards but not combinations of both[29]. It is possible that the data are not consistent with such monotonicity.

The Weibull model, as well as the exponential one, is unique in being both a proportional hazard model and an accelerated failure time model (AFT). The parameters λ and α of this distribution can be estimated by maximum likelihood procedures, controlling for censoring due to some spells being incomplete. The likelihood function can be defined as:

$$\vartheta = \prod_{Uncensored} f(t|\alpha,\lambda) \prod_{Censored} (1-F(t))(t|\alpha,\lambda) \qquad (16)$$

The first term on the right hand side is the product of the density functions for the uncensored spells, while the second represents the product of the survivor functions for the censored spells.

The parametric model associated to the distribution just introduced can be estimated by using the commercial statistical package STATA, which provides survival analysis packages for the social sciences.

Before proceeding with the model interpretation it is useful to introduce some keys concept on the hazard model estimated in this paper.

The Weibull proportional hazard (PH) model is one of the most popular parametric duration models and provides estimates for the baseline hazard and the covariate vector. The information on the nature of duration dependence is captured by the estimates for the baseline hazard and this may be quite important from a policy perspective. The baseline hazard for the Weibull specification is expressed as

$$k_2(t) = \alpha t^{\alpha-1} \qquad (17)$$

[29] The parametric model which permits non-monotonic behaviour in the hazard is the Log Logistic one. More precisely, the numerator of the hazard function is identical to the Weibull but its behaviour is more flexible because of the role of the denominator. If α < 1, the hazard decreases monotonically. If α > 1 the hazard increases from 0 at the origin to a single maximum and delineates an inverted U-shaped hazard. Thus, the hazard function in this case is non-monotonic and strongly resembles the log-normal hazard.

The most tractable functional form for the covariates is

$$k_1(x_i) = \exp(\beta' x_i) \tag{18}$$

The resulting Weibull hazard is

$$\theta_i(x;t) = \exp(\beta' x_i)\alpha t^{\alpha-1} \tag{19}$$

The choice of the functional form for the covariates ($k_1(x)$) is one of the most commonly used and this is because it renders the hazard a log-linear function of the covariates and facilitates the interpretation of the covariate effects.
In contrast to the AFT model, the effect of the covariates is on the hazard not on the time to failure (if a covariate reduces the hazard - the probability of exiting the state conditional on having been in unemployment up to the start of the analysed period - at the same time it increases the time to failure which is the time spent in this labour market status).
However, the Weibull hazard function estimates can be used to inform of the effect on duration, as many investigators are more interested in the effects of a change in the covariate on the average duration associated with a particular event (in this case it is the unemployment status). The expected duration for the Weibull hazard model can be expressed as

$$\mathrm{E}(T) = \Gamma\left(\frac{1}{\alpha}+1\right)\exp\left(-\frac{\beta' x}{\alpha}\right) \tag{20}$$

Where Γ (.) is the Gamma CDF operator.
By taking the logarithms

$$\ln(\mathrm{E(T)}) = -\frac{\beta' x}{\alpha} + \text{constant} \qquad \text{if } \alpha \neq 1 \tag{21}$$

An alternative way of expressing this in terms of log duration for the i^{th} observation is

$$\alpha \ln(T_i) = -\beta' x + u_i \tag{22}$$

Where u_i is an error term and α is the duration dependence parameter from the Weibull hazard model.
To get the marginal effect, which is the effect of the covariate on the log expected duration, it is necessary to compute the derivative as follows:

$$\frac{d\ln(E(T))}{dx} = -\frac{\beta}{\alpha} \tag{23}$$

This makes intuitive sense for a number of reasons.
First of all if the covariate x has a positive (negative) effect on the exit hazard (the hazard of exiting the unemployment state), then the covariate should exert a negative (positive) effect on the expected log duration in the state in question. For instance, a high level of education could increase the instantaneous exit rate from unemployment and thus reduce the expected (or average) unemployment duration.
Secondly, the scaling of the covariate's coefficient in this case is really relevant. If $\alpha < 1$ (or > 1), there is negative (positive) duration dependence, and the effect of a small change in the covariate on the expected duration is thus higher (lower) relative to the no duration dependence case.
It is also evident that in the case of no duration dependence ($\alpha = 1$), the conventional log linear interpretation of the estimated effect is an exponential distribution in failure time (which shows, as stated before, why the Weibull distribution is a generalisation of the exponential one).

After cleaning the data set, the analysis reported in this paper is based on 823 individuals who experienced unemployment. The dependent variable is the unemployment duration for the people who left this state[30]. More precisely, 513 individuals exited the unemployment state and it was possible to build up complete spells, while the remaining 310 were still unemployed at the time of the survey (their observations are right censored). Possible regressors include

[30] The duration of each spell of unemployment has been calculated by the author by using the information collected by the Istat's survey. More precisely, the sampled individuals were asked to provide retrospective information on their labour market experience and on their job's search actions. By joining this information it has been possible to compute the precise length of each individual's unemployment spell.

observed individual characteristics such as age, sex, geographical division, education, marital status, family type, the presence or not of child in a given household, but also economic indicators such as the local unemployment rate or the GDP level.
More precisely, the covariates employed in this analysis are primarily individual observable characteristics, classified in this way:

- ✓ indiv is a dummy variable which takes the value 1 if the individual is the head of the household, 0 otherwise. This choice comes from the fact that quite often the head is the main source of earnings of the household (this has been a peculiarity in the past).
- ✓ sex is a dummy variable taking the value 1 for a male and 0 for a female. The male component represents 41.3% of the total sample, and the female is 58.7%.
- ✓ age is the age of the individual interviewed. It ranges between 15 and 71, with a mean of almost 34 years. The range is wide because from 2004 the survey also included people aged above 64 (the 64-75 age group is now included in the people looking for a job category). However most of the sample is aged between 17 and 49.
- ✓ mstatus: marital status is a dummy variable taking the value 1 if the individual is married and 0 otherwise (single, divorced, separated legally or not, widowed).
- ✓ child: dummy variable taking the value 1 if the household (married or not, as defined by the survey) has child, 0 otherwise.
- ✓ URate: this is the quarterly unemployment rate, obviously computed for each of the 8 quarters covered by the data.
- ✓ married: dummy variable taking the value 1 if the individual has been married at least once (including married now, divorced, separated legally or not, widowed), 0 otherwise. The decision to include an additional variable for the marital status was because someone married or who has been married has more economical

constraints than a single person. This is due to Italian law by which often it is the head of a household (married now or previously) who has to provide money for child care and education and is also obliged to support his ex-wife (often classified as the weak part of the couple, because she has to take care of the child and she is often unemployed/out of the labour force for this reason).

- ✓ dummy variables for the provinces' geographical location. 4 dummies were generated: north if the province was located in the north of Lombardy (including Como, Sondrio, Bergamo, and Lecco), south for the provinces of Cremona, Mantova and Lodi, west for Varese, Milan and Pavia, and east for the province of Brescia.

To avoid the collinearity problem it was necessary to select a base category to be excluded from the covariates' list. In this case the choice was not driven by economic criteria, because there are no substantial differences between these cities. Given that the east dummy includes only one province it was chosen as base for interpretation of the results.

- ✓ The categorical variable for education was collapsed into two dummy variables. The first one (educ1) is one if people attended middle school (implying on average eight years of schooling), and 0 otherwise. The second one is one for people who received higher education (including diploma and degree, therefore at least 13 years of schooling).

  As for the provinces, one dummy was selected as base category to avoid collinearity between the covariates. In this case the variable excluded is educ1.

3.3 The estimates

The model estimates were obtained by using the Stata statistical package, which provides a specific range of commands to analyse the survival time data[31].

First of all it is necessary to define the data as being survival–time. This leads to the definition of the dependent variable which is spell. As explained before, it is the duration of the unemployment spell for each person in the sample. The failure event occurs if the individual exits the state of unemployment. In this model this is captured by the dummy variable newunemp, which is 1 if the individual is still unemployed at the time of the survey, 0 otherwise (there are two possible alternative outcomes: to be employed (dummy variable newemp) or to go out of the labour force (dummy variable newolf). The right censored observations are equal to the number of people still unemployed when interviewed.

As stated before, the total number of individuals in the sample was 823 (selected from the original sample as the sub sample of people who experienced unemployment). There were 310 people still unemployed at the time of the survey, and this is the number of right censored observations. Therefore, 513 individual left the unemployment state and information about their actual labour market status was collected.

[31] The commands for the survival time analysis and their explanations are described in the basic Stata manual *Stata base reference manual. V.4, S-Z* (2003), College Station, Tex, Stata Corporation

The last step before starting the model estimation is to declare what covariates are going to be included in the estimation. As described before, all the regressors will be kept with the exception of the base categories for the provinces' dummies (east which corresponds to the province of Brescia) and for the education (educ1, corresponding to middle school level).

In order to capture additional insights on the variable employed it is useful to compute the summary statistics which, as known, give information on the number of observations, the mean, the coefficient of variation, and the range within which each variable is defined.

Table 3: Summary statistics of the variables used in the analysis

Variable Name	Obs	Mean	CV #	Min	Max
spell	823	20.1762	2.1673	1	417
indiv	823	0.2260	0.8518	0	1
sex	823	0.4131	1.1927	0	1
age	823	33.9222	0.3528	15	71
mstatus	823	0.3694	1.3072	0	1
child	823	0.8347	0.4452	0	1
URate	823	4.0226	0.0518	3.7	4.5
married	823	0.4544	0.4982	0	1
north	823	0.2795	1.0964	0	1
east	823	0.0826	3.3353	0	1
south	823	0.2089	1.9473	0	1
west	823	0.4289	1.1546	0	1
educ1	823	0.4678	1.0673	0	1
educ2	823	0.5322	0.9382	0	1
newunemp	823	0.3767	1.2870	0	1
newemp	823	0.2770	1.6166	0	1
newolf	823	0.3462	1.3752	0	1

Note: # CV is the coefficient of variation. It is defined as the ratio of the standard deviation to the mean

Source: Elaborations from Stata

A detailed description of the dependent variable (spell) and of all the regressors included in the model has been given above. As regards the last three variables listed in the above table, it is convenient to introduce a proper explanation of their meanings. Newunemp is a dummy variable of 1 if the individual is still joblessness at the time of the survey and 0 otherwise. Newemp is another dummy variable of 1 if the interviewee that experienced unemployment in the past exited this state successfully. The last dummy (newolf) is 1 if the ex-unemployed is now out of the labour force.

These three variables will be used, together with another one[32], to analyse four different competing risks.

The model summarized is shown in the following table

Table 4: Unemployment Duration: Weibull model's estimates

Explanatory Variable	**Coef. (std error)**	**z**	**P>│z│**	**[95% Conf.**	**Interval]**
indiv	0.137 (0.167)	0.82	0.412	-0.190	0.465
sex	0.384 (0.127)	3.01	0.003	0.134	0.634
age	-0.034 (0.007)	-4.53	0.000	-0.049	-0.019
mstatus	-0.637 (0.212)	-3.01	0.003	-1.052	-0.222
child	-0.195 (0.159)	-1.23	0.220	-0.508	0.117
URate	0.051 (0.276)	0.19	0.853	-0.490	0.592
married	0.667 (0.246)	2.71	0.007	0.184	1.149
north	-0.599 (0.220)	-2.72	0.006	-1.030	-0.167
south	-0.486 (0.224)	-2.17	0.030	-0.925	-0.048
west	-0.658 (0.209)	-3.14	0.002	-1.069	-0.247
educ2	0.497 (0.118)	4.20	0.000	0.265	0.729
_cons	-1.922 (1.205)	-1.59	0.111	-4.284	0.440
α	0.782 (0.031)			0.722	0.846
Obs	823				
LogL	-816.42				

Source: Elaborations from Stata

It can be inferred that the covariates used to explain joblessness duration are indiv, sex, age, marital status, child, URate, married, north, south, west and educ2. It is important to remember that the base categories for the provincial

[32] As it will be clarified below, the fourth risk is represented by the dummy variable dipind, that takes the value one if the individual after an unemployment experience find a dependent job and 0 if he is self-employed. All these four dummies will be renamed as risk 1, risk 2, risk 3, and risk 4 for the CRM estimation.

dummy and the education ones are east (Brescia) and educ1 (up to primary school) respectively.

As explained above, a small change in the value of the regressor will induce a proportional change in the hazard rate, given the logarithmic transformation used.

In regard to the binary variables, again given the logarithmic nature of the dependent variable, it is required to compute [exp[β_k - 1] · 100 to yield the effect on the hazard rate in percentage terms. By proceeding to interpret the reported estimates, the dummy variable for the head of the household does not exert a statistically significant impact on the unemployment exit hazard rate. However, the sign of this covariate suggests that, as expected, being head of the household should increase the hazard rate by reducing the time spent in unemployment with respect to the other household components.

Sex is significant at the 1% level, and the sign of the variable suggests that being male, on average and ceteris paribus, increases the hazard rate by [exp(0.38) - 1] · 100 = 46.23% with respect to a woman with the same characteristics. Also age is significant at 1% level and with the expected sign: getting older by one year reduces the hazard of exiting the state by [exp(-0.034) - 1] · 100 = -3.4%. As unemployed people get older they increase their permanence in this state.

Being married now, on average and keeping the other covariates constant, reduces the probability of leaving the state by [exp(-0.64) - 1] · 100 = -47,3%. This effect has apparently no intuitive sense, but it is probably due to the fact that married people (for which the dummy variable is 1) does not include those previously married (i.e. divorced, separated, widowed). The differences obtained by grouping these categories differently are shown by the inclusion of the binary variable married. It is 1 if the person has been married at least once (including in this way not only married people but also the divorced, widowed and separated, legally or not) and 0 otherwise (single). In this case the coefficient is still significant at 1% and takes the expected sign. The hazard rate for the "married at least once" is, on average and ceteris paribus, [exp(0.66) - 1] · 100 = 93.5% higher than for the singles. This result is quite obvious and likely comes from the presence of economic constraints also for people (typically the ex-head of the household) previously married, who have to provide for child care and education (sometimes these constraints include providing for an ex wife, if she is defined as the weakest part of the household).

Having a child exerts no impact on the exit hazard rate. Furthermore the sign of this binary variable seems counterintuitive, because the hazard for couples with a child is smaller than for households without a child. However this effect could be strongly influenced by the wife's behaviour, given that quite often a woman with a child leaves her employment and remains unemployed at least until the child starts middle school (on average at the age of 6). Also the unemployment rate does not exert a significant impact on the hazard and the sign is not the expected one. This is because an increase in the unemployment rate (i.e. by one percent) would raise the hazard of leaving this state at the same time reducing the time to failure.

The province variables are all statistically significant and all take the same sign.

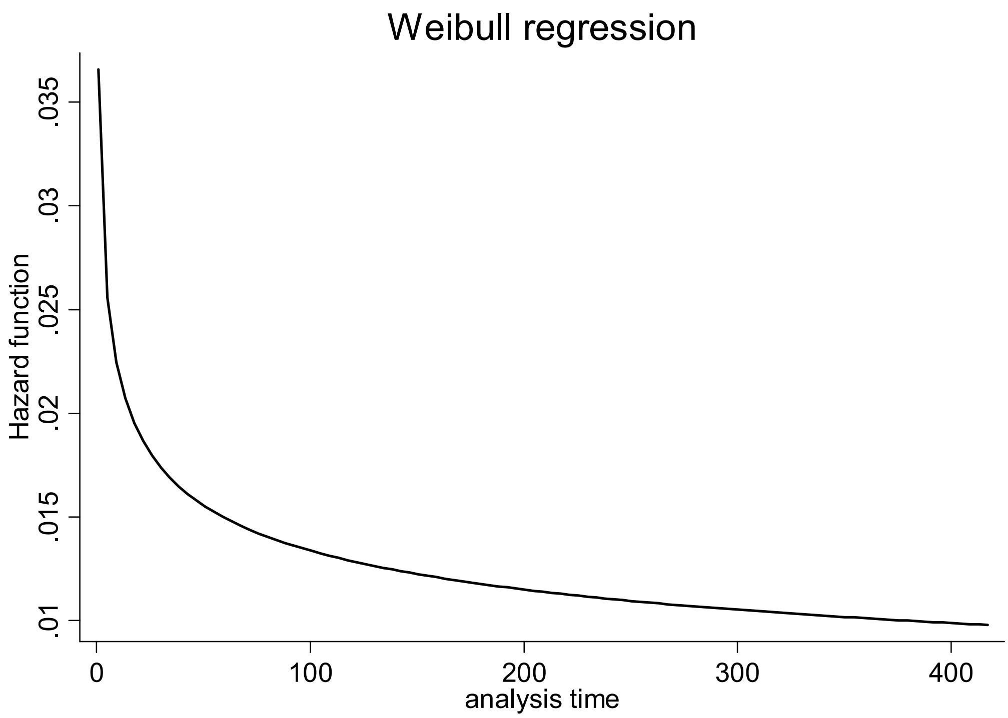
The hazard rate for an individual living in the north (significance level of 1%) of Lombardy is, on average and ceteris paribus, $[\exp(-0.59) - 1] \cdot 100 = -44,6\%$ lower than for people living in Brescia (east, base category). The hazard reduction for an individual living in the south and in the west are $[\exp(-0.48) - 1] \cdot 100 = -38.1\%$ and $[\exp(-0.65) - 1] \cdot 100 = -47.8\%$ respectively, with a significance level of 5% and 1%. These effects suggest that Brescia is the province with the fewest unemployment problems.

The last variable summarizes the effect of education on the probability of leaving unemployment. It is significant (1% level) and with the expected sign. More precisely, the hazard for individuals holding a high qualification (diploma, degree or above) is $[\exp(0.49) - 1] \cdot 100 = 63.2\%$ higher than for people educated to middle school level.

It is useful to summarize the main implications of the model estimated. Males and married (at least once) individuals have shorter expected durations relative to females and singles, and the results are robust. Also getting older reduces the hazard of leaving unemployment. The result that an educational level higher than middle school (which implies eight years of schooling) contributes to higher probability of exiting is both plausible and consistent with past studies[33] . The province's indicators suggest that Brescia (classified as east) is the city with the fewest unemployment problems.

The estimated coefficient for the duration dependence parameter suggests the presence of negative duration dependence as well as of a monotonically decreasing Weibull distribution. This effect becomes clear by graphing the baseline hazard function. In this way we can notice a reduction of the probability of exiting (the predicted hazard function) as the time spent in unemployment increases.

Figure 4: Estimated Baseline Hazard Function after parametric regression for the Weibull distribution



Source: Elaborations from Stata

An asymptotic t-test to test the proposition of no duration dependence, H_0: α=1 versus H_a: α≠1 yields an asymptotic t-value of -6.95. The critical value is ± 1.96. Thus, the null hypothesis of constant duration dependence is decisively rejected

[33] In their work, Trivedi and Alexander (1989) stress this issue.

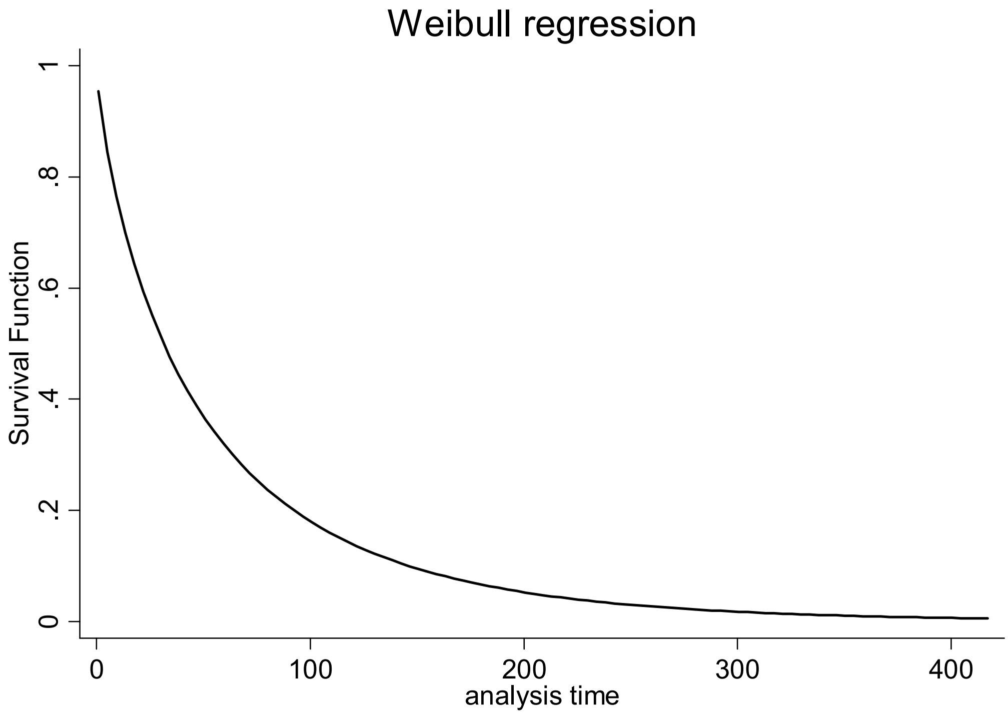
by these data. It is clear that there is strong negative duration dependence. Thus, the longer a spell of unemployment has gone on, the less the likelihood it will be left.

One of the biggest issues for the duration analysis is that of neglected heterogeneity. Allowing for this may explain any role for unobservable characteristics that have not been controlled within the regression model. If there is and there is no control for it, the inference would be flawed (often leading to an overestimation of the duration dependence parameter). In this case, there is no change for the estimated coefficients for the covariates and the duration dependence parameter maintains its absolute value[34].

The associated survivor function for the Weibull proportional hazard (PH) model, evaluated at the means of the explanatory variables, is illustrated in the next figure.

Figure 5: Estimated Baseline Survival Functions after parametric regression for the Weibull distribution

[34] As briefly stated, unobserved heterogeneity does not change the estimates, and also the duration dependence parameter shows again evidence of negative duration dependence. I know that I have to be careful assigning time dependence, but I have also to take into consideration the data constraints. Those mainly refer to the time span covered, that is quite short (8 quarters data for the years 2004 and 2005), and maybe too short for a precise duration dependence identification. Again, this could affect also the neglected heterogeneity behaviour. I think that this issue, deeply discussed in Arulampalam (2001 and 2002), Arulampalam, Booth and Taylor (2000), Arulampalam and Stewart (1995), does not change my results also because of the nature of the covariates employed (mainly individual/observed characteristics). The precision of the estimates will improve with the Italian data that will cover a wider time span and obviously will include a larger number of unemployed.



Source: Elaborations from Stata

Where, clearly, survival is 1 at the start of the time analysis, quickly declines up to 100 months of duration, showing thereafter a steady reduction.

Table 5: Marginal effects after the Weibull regression

Variable	dy/dx	Std.Error	z	P>│z│	[95% Conf. Interval]		X
indiv*	-5.2669	6.1351	-0.86	0.391	-17.2915	6.7576	0.2260
sex*	-14.9538	4.9699	-3.01	0.003	-24.6947	-5.2128	0.4131
age	1.3683	0.3117	4.39	0.000	0.7574	1.9792	33.9222
mstatus*	29.3131	11.6881	2.51	0.012	6.4046	52.2217	0.3693
child*	7.2513	5.4945	1.32	0.187	-3.5178	18.0205	0.8347
URate	-2.0616	11.1331	-0.19	0.853	-23.8823	19.7591	4.0226
married*	-26.5872	10.2931	-2.58	0.010	-46.7619	-6.4124	0.4544
north*	29.2334	13.3121	2.20	0.028	3.1414	55.3253	0.2795
south*	23.8296	13.3111	1.79	0.073	-2.2596	49.9188	0.2089
west*	28.9526	10.5241	2.75	0.006	8.3266	49.5786	0.4289
educ2*	-20.7793	5.5031	-3.78	0.000	-31.5653	-9.9933	0.5322

(*) dy/dx is for discrete change of dummy variable from 0 to 1
Source: Elaborations from Stata

The marginal effect, as mentioned above, explains the impact of an infinitesimal change of a covariate on the dependent variable. For the survival data analysis Stata reports the following

$$\partial \ln(E(T))/\partial X = - \beta/\alpha \tag{24}$$

This is the derivative of the log duration (called spell in this estimation) with respect to each covariate. As shown on the right hand side this corresponds to the negative of the ratio of the model coefficient (the Weibull one estimated before) on the duration dependence parameter (α). In this way, clearly, the estimated coefficients will assume the opposite sign with respect to the model ones. For this reason their effect is on the duration (and not in terms of hazard rate). Furthermore, given the negative nature of the duration dependence they will increase in absolute value. In particular it can be inferred that the coefficients significant in the Weibull estimation obviously maintain their relevance. However, for some of the dummy variables there can be inferred an increase in the significance level.

4 Competing Risks

Before presenting the empirical findings, let's look at some general concepts that are used in the competing risks model (CRM). Often these are extensions of concepts already mentioned for the duration analysis.

The basic CRM formulation is applicable to modelling time in one state when exit is to a number of competing states, such as different labour market states.

The CRM analysis is really useful because it has the potential to provide information on issues currently of great interest to economists, and especially useful from a policy perspective. As will be shown below, one valuable application for the purpose just listed is labour market data. Further, the CRM is relatively straightforward to implement if the model is a Proportional Hazard (PH)[35].

We consider the CRM in which there are *m* latent duration or failure times, one for each distinct competing cause of failure. The setup of the model is the following.

Each subject has an underlying failure time, which is subject to censoring.

Failure time may be one of *m* different types, given by the set $J = (1,...,m)$. We may think of this as a situation with *m* different causes of transition from a given state (in this paper it is the unemployment). However, the occurrence of a failure of one kind of event removes the individual from risks of other kinds of events. Therefore, given censoring of the remaining (*m* – 1) durations for each individual, we observe at most one complete duration.

In a CRM with *m* types of failures, there are *m* + 1 states {0, 1, …, *m*}, where 0 represents the initial state and {1, …, *m*} are possible destination states. For the *i*th individual the data vector is of the form $(x_i, t_i, d_{1i}, ..., d_{mi}, d_{ci})$, where x_i is a vector of weakly exogenous covariates that measure the characteristics of *i*, $t_i = \min(t_{1i}, ..., t_{mi}, t_{ci})$, where t_{ki} denotes the time to transition to the *k*th destination, t_{ci} denotes the time of censoring, and $d_{ji} \equiv 1(t_{ji} = t_i)$, j = 1,…, *m*, *c* are dummy variables that take the value of $t_{ji} = t_i$. Because we only observe one of the t_{ji}, the remaining are interpreted as latent variables.

[35] A description of this model has been provided in section 3.

Censoring may be regarded as a competing risk. It operates on individuals according to a probability distribution. In this paper, as will be explained below, the censoring variable is assumed to be independent of the $(t_1,...,t_m)$.

Since allowing for neglected heterogeneity does not lead to a change in the estimation results (in particular the duration dependence parameter maintains the same value), also for these estimations we do not include this issue. More precisely, for each of the hazard functions estimated we constrain the slope coefficient to be independent of the state of origin[36].

CRM are often used as an extension to duration models, because they can inform on the destination states of the individuals analysed. In our application CR model is used to delineate the competing causes for exiting from unemployment. Consider an individual who experienced unemployment and is "at risk" of transiting to employment, or remaining unemployed, or leaving the labour force, or getting a dependent job. Succumbing to any one condition means that transition to the other states is not possible. In fact in an *m*-event setup, each event provides one complete duration and *m* – 1 censored durations. Thus we have a situation of competing risks in which there is competition to determine the unemployed individual's destination state.

Although discrete-time models are often required in empirical applications, our analysis of the joint hazard formulation uses the continuous-time framework and follows the exposition given in Mealli and Pudney (1996). Furthermore we assume that we have single-spell data[37].

The model provides the joint distribution of the spell duration, denoted τ, and the exit route *r*, which is an integer variable that takes one of the values in the set $(1,2,...m)$.

We ignore censoring for simplicity and assume that there exist latent variables $(t_1,...,t_m)$, one for each destination, that correspond to the spell duration for each possible exit route by which the spell may terminate if there were no other

[36] For more details on the duration dependence, see Arulampalam (2001 and 2002), Arulampalam, Booth and Taylor (2000), Arulampalam and Stewart (1995).

[37] These assumptions are mainly driven by the kind of data available and by the consequent analysis which these statistics enable. However these assumptions are quite common in many empirical studies and also in econometrics books. For further details, see Cameron and Trivedi (2005).

risk factors that might cause the spell to end sooner. Destination-specific covariates are denoted by $x_j(j=1,...,m)$. We observe one duration, *t*, where

$$\begin{aligned}\tau &= \min(t_1,...,t_m) \\ &= \min_j(t_j), t_j > 0\end{aligned} \tag{25}$$

at the termination of the spell; that is, only the shortest duration is observed and the rest are censored. Censoring owing to factors other than exit are not considered. Then

$$\begin{aligned}\Pr[\tau > t] &= \Pr[t_1 > t,...,t_m > t] \\ &= S_\tau(t)\end{aligned} \tag{26}$$

is the joint survivor function. If the risks are independent then

$$\Pr[\tau > t] = \Pr[t_1 > t] \times \Pr[t_2 > t] \times ... \times \Pr[t_m > t]. \tag{27}$$

The corresponding exit route *r* is given by

$$r = \arg\min_{j \in J}(t_j) \tag{28}$$

Let $g_j(t)dt$ denote the probability of entering into state (risk) *j* in the interval $(t, t+dt)$; then the total hazard rate applicable to all causes is

$$\lambda_\tau(t) \equiv -d/dt \ln S_\tau(t) = \sum_{j=1}^{m} g_j(t). \tag{29}$$

If risks are independent, then the hazard rate for a specific cause *j* is $\lambda_j(t) = g_j(t)$. This means that probability of failure from cause *j* in $(t, t+dt)$, conditional on survival to *t*, is the same whether *j* is one of the risks or the only risk.

The probability of surviving the risk *j* in the interval (T_1,T_2) conditional on survival to T_1 is

$$\begin{aligned}\int_{T_1}^{T_2} \lambda_j(t)dt &= \int_0^{T_2} \lambda_j(t)dt - \int_0^{T_1} \lambda_j(t)dt \frac{\partial^2 \Omega}{\partial v^2} \\ &= \ln S(T_2) - \ln S(T_1) \\ &= \ln \frac{Pr[t_j > T_2]}{Pr[t_j > T_1]},\end{aligned} \tag{30}$$

or equivalently

$$\exp\left(-\int_{T_1}^{T_2} \lambda_j(t)dt)\right) = \frac{\Pr[t_j > T_2]}{\Pr[t_j > T_1]}. \tag{31}$$

One minus the left-hand side expression is referred to as the net probability of leaving unemployment for cause *j* in the interval (T_1, T_2).

We can now include in the picture covariates that affect the hazard rate. By assuming, as mentioned above, independent risks (as opposed to correlated risks) and considering the distribution of t_j, the hazard rate for failure of *j*th type is defined by

$$\lambda_j(t_j|x_j) = \lim_{\Delta t_j \to 0} \frac{\Pr[t_j \le T \le t_j + \Delta t_j T \ge t_j, x_j]}{\Delta t_j} \tag{32}$$

and the integrated hazard $\Lambda_j(t_j|x_j)$ for the *j*th type risk is

$$\Lambda_j(t_j|x_j) = \int_0^{t_j} \lambda_j(s|x_j)ds. \tag{33}$$

Then the duration density is

$$\begin{aligned} f_j(t_j|\mathrm{x_j}, \beta_j) &= \lambda_j(t_j|\mathrm{x_j}, \beta_j) S_j(t_j|\mathrm{x_j}, \beta_j), \\ &= \lambda_j(t_j|\mathrm{x_j}, \beta_j)\exp[-\Lambda_j(t_j|\mathrm{x_j}, \beta_j)],\end{aligned} \tag{34}$$

Using the relationship between survivor and integrated hazard functions. Defining $x = [x_1, ..., x_m]$ and $\beta = [\beta_1, ..., \beta_m]$ gives the joint density of τ and *r*:

$$
\begin{aligned}
f_j(\tau, r|x, \beta) = {} & f_r(\tau|x_r, \beta_r)\prod_{j\neq r} e\mathrm{xp}[-\Lambda_j(\tau|x_j, \beta_j)] \\
& = \lambda_r(\tau|x_r, \beta_r)\exp[-\Lambda_r(\tau|x_r, \beta_r)] \\
& \quad \times\prod_{j\neq r} e\mathrm{xp}[-\Lambda_j(\tau|x_j, \beta_j)] \\
& = \lambda_r(\tau|x_r, \beta_r)\prod_{j=1}^{m}\exp[-\Lambda_j(\tau|x_j, \beta_j)].
\end{aligned} \tag{35}
$$

The first line follows from the product of conditional and marginal probabilities. The second term on the right-hand side is the product of survival probabilities for all exit routes other than *r*, which uses the independence of risks assumption. Expression (35) implies that

$$
\begin{aligned}
& \lambda_j(\tau|x_j, \beta_j)\exp\left[\sum_{j=1}^{m}-\Lambda_j(\tau|x_j, \beta_j)\right] \\
& = \lambda_j(\tau|x_j, \beta_j)\exp\left[-\Lambda^a(\tau|x, \beta)\right],
\end{aligned} \tag{36}
$$

where $-\Lambda^a(\tau|x, \beta) = \sum_{j=1}^{m}\Lambda_j(\tau|x_j, \beta_j)$ is the aggregate or overall integrated hazard. This equation simply says that the total hazard of leaving the origin state is the sum of hazards for all destinations. The overall survivor function is

$$
S(t) = \exp(-\Lambda^a(t)). \tag{37}
$$

The likelihood function given independent risks is the product over all observations of terms like (34). This likelihood can be written explicitly if all relevant functional forms are specified. Many issues relevant also for the duration analysis, such as flexibility of functional form, unobserved heterogeneity, and so forth, remain relevant in the context of CRM.

Before proceeding with the empirical evidences it is worth underlining that the parametric CR Weibull model is the one that best fits the data also for the kind of analysis just introduced. A detailed description of this distribution has been provided in section 3.

4.1 What is the destination state of Lombardy's unemployed?

4.1.1 Estimates under Competing Risks Framework

The duration analysis presented in section 3 focused on the time in an unemployment spell, ignoring the destination state after transition. Here we implement competing risks analysis of the same data used for the survival one. As explained above, we selected a sub sample which covered 823 individuals experiencing unemployment (each corresponding to an individual spell), 310 of which still unemployed when interviewed. The data identify four different destination states: exit with success from unemployment by reaching the employment state (risk 1), either employee or self-employed after the unemployment experience (risk 2), still jobless (right censored observations, corresponding to the 310 individuals mentioned, risk 3), and leave the labour force (risk 4). Before describing the models it is worth relaxing the assumption that the hazard function does not depend on the destination state and to consider instead the competing risks formulation.

For Lombardy's data set there are 228, 172, 310, and 285 transitions, respectively, to each of the four states mentioned. For each transition we estimated four parametric models. More precisely, Weibull, exponential, Cox PH and Log Logistic model were evaluated, without allowing for neglected heterogeneity given that, as explained above, it does not affect the estimates. Since the CR Weibull model is the one that best fits our data, the analysis will be focused on these results, while the other models' estimates are shown in table A3 (appendix). Because of the assumption of independent competing risks, estimation can be carried out one equation at a time. The variables are summarized in table 6 for the sample we use for estimation and selected extracts of the computer output, with focus on all the covariates employed, are given in table 7.

Table 6: Sample means and standard deviations of variables by spell type (2004-2005)

	Employment (Risk 1)		Employees (Risk 2)		Unemployment (Risk 3)		Out of the labour force (Risk 4)	
No of episodes	228		172		310		285	
Variable	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev
indiv	0.2412	0.4289	0.2151	0.4121	0.2419	0.4289	0.1965	0.3980
sex	0.5570	0.4978	0.5465	0.4993	0.4226	0.4948	0.2877	0.4535
age	31.8289	11.1774	31.7151	10.9563	33.7452	11.6598	35.7895	12.6393
mstatus	0.2939	0.4565	0.2907	0.4554	0.3355	0.4729	0.4667	0.4998
child	0.8465	0.3613	0.8663	0.3413	0.8225	0.3826	0.8386	0.3685
URate	4.0202	0.2198	4.0401	0.2199	4.0406	0.2089	4.0049	0.1974
married	0.3640	0.4822	0.3605	0.4815	0.4452	0.4978	0.5368	0.4995
north	0.3026	0.4822	0.2907	0.4554	0.2516	0.4346	0.2912	0.4551
south	0.1535	0.3613	0.1686	0.3754	0.2355	0.4250	0.2245	0.4180
west	0.4561	0.4992	0.4302	0.4965	0.4193	0.4042	0.4175	0.4940
educ2	0.5746	0.4955	0.5523	0.4987	0.5677	0.4962	0.4596	0.4992

Source: Elaborations from Stata

Table 7: Estimates of the Weibull CR model (standard errors)

	Risk 1	Risk 2	Risk 3	Risk 4
Transitions	228	172	310	285
Variable				
indiv	0.358	0.174	0.137	-0.064
	(0.202)	(0.239)	(0.167)	(0.183)
sex	0.742	0.713	0.384	-0.263
	(0.147)**	(0.170)**	(0.127)**	(0.147)
age	-0.045	-0.038	-0.034	-0.015
	(0.009)**	(0.011)**	(0.008)**	(0.007)*
mstatus	-0.236	-0.272	-0.637	0.247
	(0.288)	(0.333)	(0.212)**	(0.257)
child	-0.041	0.094	-0.195	-0.019
	(0.199)	(0.240)	(0.160)	(0.171)
URate	-0.400	0.086	0.051	-0.711
	(0.324)	(0.366)	(0.276)	(0.294)*
married	0.327	0.317	0.667	-0.143
	(0.328)	(0.380)	(0.246)**	(0.291)
north	-0.154	-0.409	-0.599	-0.171
	(0.255)	(0.271)	(0.220)**	(0.258)
south	-0.630	-0.730	-0.486	-0.261
	(0.283)*	(0.299)*	(0.224)*	(0.265)
west	-0.304	-0.567	-0.658	-0.374
	(0.247)	(0.261)*	(0.210)**	(0.252)
educ2	0.415	0.297	0.498	-0.052
	(0.137)**	(0.157)	(0.118)**	(0.122)
_cons	-0.001	-2.176	-1.922	0.691
	(1.414)	(1.607)	(1.205)	(1.277)
α	0.588	0.567	0.782	0.748

	(0.031)	(0.034)	(0.031)	(0.031)
Obs	823	823	823	823
LogL	-778.96	-647.16	-816.42	-774.63

Notes: (i) Standard errors in brackets.
(ii) * ** denote significance at 5% and 1% respectively.
(iii) All results are obtained using STATA.

Before proceeding with a detailed interpretation and comparison of the CR estimates obtained it is important to stress that the Weibull model specified for risk 2 (employee or self-employed after the joblessness) has the highest log-likelihood, -647.16. Then we find the models estimated for risk 4 and risk 1, with log-likelihood of -774.63 and -778.96 respectively. For individuals still jobless we obtain the lowest log-likelihood (-816.42), but it is the model which provides the highest number of significant coefficients (mainly at 1% and 5% significance level). We can now better discuss the results for the competing causes analysed. As is shown in table 7, the coefficients of indiv remain imprecisely determined for all the four competing causes, but the sign is as expected. While being head of the household (indiv=1) accelerates transitions out of unemployment of those seeking a job (risk 1 and risk 2) and increases the hazard function for people without a job (risk 3), it reduces the transitions from unemployment to out of the labour force, with respect to the other family members.

The coefficients of sex are significant for all the risks, and positive in risk 1, risk 2, and risk 3, but not in the fourth. That is, males have high employment probabilities and a higher hazard for the third risk, but lower hazard of leaving unemployment by exiting the labour force with respect to a woman with the same characteristics. It is worth underlining the presence of a discouragement effect for female labour force participation.

The estimates of age are significant and negative in all the hazard functions. Getting older reduces transitions out of unemployment of those seeking employment, of the jobless, and of individuals leaving the labour force. The impact of married and mstatus is significant for the third risk but it is negligible for the other labour market transitions. More precisely, being married now (mstatus) or at least once (married) decelerates transitions out of unemployment of those seeking a job, but accelerates the movements of people leaving the labour force with respect to single people. The presence of a child remains imprecisely determined for all the transitions. For the URate, it is necessary to underline that an increase in this indicator decelerates transition out of the labour force for individuals previously unemployed (with a significance level of 5%).

The dummy variables summarizing the geographical partitions suggest that, living in the north, south, or west of Lombardy leads to a reduction in the hazard of exiting the labour force after unemployment with respect to living in Brescia (east, base category).

The coefficients of educ2 are positive for all the risks except the fourth. As expected, a higher level of education accelerates the probabilities of exiting with success from unemployment (by finding a job), and reduces the likelihood of leaving the labour force after unemployment experience. Furthermore, the coefficients for these dummy variables are significant for all risks, with the exception of risk 2.

The main implications of the CRM estimated - useful from a policy perspective - are now summarized. Males have relatively higher employment probabilities and shorter unemployment duration (a result also shown in the previous analysis) with respect to females. Furthermore the male component of the labour force has a lower likelihood of leaving the labour force after unemployment experience, again with respect to females. This result strongly suggests (significance level) the existence of a discouraged workers effect for the female component. More precisely, this effect leads to a negative correlation between the unemployment rate and the labour force participation rate. As a result, females are more likely to become inactive after a period of joblessness. However this behaviour could also be driven by other relevant factors. One of the most important is childcare. Quite often, married women leave the labour force to take care of their children. This has been a very strong issue for OECD countries, but new evidence confirms that policies to promote female labour force participation will be introduced[38].

Getting older reduces all the transitions analysed, increasing in this way unemployment durations. Marital status and the presence of a child do not exert a significant impact on the transitions analysed. The east of the region is the area with the least unemployment problems. In fact, as mentioned for the duration analysis, living in Brescia increases the hazard of leaving unemployment. Holding higher educational qualifications than middle school one (which implies eight years of schooling) contributes to a higher probability of exiting with success from the unemployment state and reduces transitions out of the labour force. This result, consistent with previous studies, emphasizes the role of education in the reduction of unemployment.

5 Conclusions

The aim of this paper was to analyse unemployment duration and the CRM for Lombardy's labour market. These methodologies are becoming increasingly pervasive in applied research to explain factors determining both the time spent in a state (in this case unemployment) and the exit route from this state (leaving unemployment with success or exiting the labour force).

After presenting the European context in section 2 and the main empirical evidence on unemployment, section 3 gives a detailed description of the data used and an empirical application of the unemployment duration analysis to Lombardy's labour market. After a precise description of the CR technique, section 4 presents the empirical evidence and the implications for the Regional context analysed in this paper.

It is worth underlining the main findings of the analyses applied, since these provide significant insights for policy analysis.

Males and married people (at least once) are the individuals which experience shorter joblessness durations relative to females and single people, and the results are robust. Also getting older reduces the hazard of leaving

[38] For further details, see OECD (2004).

unemployment. The role of education is confirmed in this study, since people who are better educated (with an education level higher than the middle school) have a greater likelihood of exiting unemployment. The east of the region (Brescia) is the area with the least unemployment problems.

The findings of the CR analysis are often extensions of that for duration. We find again that males have shorter joblessness duration, but we can add that they also have higher employment probabilities with respect to females. Furthermore males have a lower likelihood of leaving the labour force after an unemployment experience with respect to females. This result strongly suggests the existence of a discouragement effect for females. Getting older reduces all the labour market transitions analysed, increasing in this way unemployment duration and confirming the findings of the first analysis. The role of education in the reduction of unemployment is confirmed for the competing causes examined.

The relevance of these analyses is one of the reasons that suggest the opportunity to add further research on these issues. This objective will be satisfied by applying the methodologies used in this paper to the Italian labour market.

Bibliography

Akerlof, G. A. and Main B. G. M. (1980), *Unemployment Spells and Unemployment Experience*, The American Economic Review, Vol. 70, No 5, pp. 885-893.

Arulampalam, W. (2001), *Is unemployment really scarring? Effects of unemployment experiences on wages*, mimeo, University of Warwick Department of Economics.

Arulampalam, W. (October 2002), S*tate dependence in unemployment incidence: evidence for British men revisited*, IZA Discussion Paper No 630 (Revised Nov 2004).

Arulampalam W., Booth A.L. and Taylor M.P. (2000), *Unemployment persistence*, Oxford Economic Papers, Vol.52, pp. 24-50.

Arulampalam, W. and Stewart, M. B. (1995), *The determinants of individual unemployment durations in an era of high unemployment*, Economic Journal, Vol. 105, pp. 321-332.

Baker, G. M. and Trivedi, P. K. (1985), *Estimation of Unemployment Duration from Grouped Data: A Comparative Study*, Journal of Labor Economics, Vol. 3, No 2, 153-174.

Bergström, R. and Edin, P. A. (1992), *time Aggregation and the Distributional Shape of Unemployment Duration*, Journal of Applied Econometrics, Vol. 7, No 1, pp. 5-30.

Butles, J. S., Anderson, K. H. and Burkhauser (1989), *Work and Health after Retirement: A Competing risks model with Semiparametric Unobserved Heterogeneity,* Review of Economics and Statistics, Vol 71, pp. 46-53.

Cameron, A.C. and Hall, A.D (2003), *A Survival Analysis of Australian Equity Mutual Funds*, Australian Journal of Management, Vol. 28, No 2, pp. 209-226.

Cameron, A.C. and Trivedi, P.K (2005), Microeconometrics: Methods and Applications, Cam-
bridge University Press.

Clark, K. and Summers, L. (1979), *Labor Market Dynamics and Unemployment: A Reconsideration*, Brookings Papers on Economic Activity, No 1, pp. 13-60.

Contini, B. and Trivellato, U. (2005), *Eppur si muove. Dinamiche e persistenze nel mercato del lavoro italiano*, il Mulino (Bologna).

Greene, W.G. (2003), Econometric Analysis, 5th edition, Prentice-Hall.

Han, A. and Hausman, J. A. (1990), *Flexible Parametric Estimation of Duration and Competing Risk Models*, Journal of Applied Econometrics, Vol. 5, No. 1, pp. 1-28.

Istat (Istituto Nazionale di Statistica, 2006), *La rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione*, Metodi e Norme n. 32, Roma.

Istat (Istituto Nazionale di Statistica, 2006), R*apporto Annuale. La situazione del Paese nel 2005*, Litosud, Roma.

Karr, W. (1997), *Conceptual Problems in the Understatement of Long-Term Unemployment*, Labour Market Research Topics no 21, Nürnberg, Iab.

Kiefer, N.M (1988), *Economic Duration Data and Hazard Functions*, Journal of Economic Literature, Vol. 26, no. 2, pp. 646-679.

Lancaster, T. (1979), *Econometric Methods for the Duration of Unemployment*, Econometrica, Vol. 47, No 4, pp. 939-956.

Lancaster, T. (1992), *The Econometric Analysis of Transition Data*, Cambridge University Press.

Machin, S. and Manning, A. (1998), *The Causes and Consequences of Long-Term Unemployment in Europe*, in Ashenfelter, O. and Card, D. (1999), *Handbook of labor Economics*, Vol. 3C, pp.3085-3139.

Maddala, G.S. (1983), *Limited-Dependent and Qualitative Variables in Economics*, Cambridge University Press.

McCall, B. P. (1996), *Unemployment insurance rules, joblessness and part-time work*, Econometrica, Vol. 64, No 3, pp. 647-682.

Mealli, F. and Pudney, S. (1996), *Occupational Pensions and Job Mobility in Britain: Estimation of a Random-Effects Competing Risks Model*, Journal of Applied Econometrics, Vol. 11, No 3, pp. 293-320.

Mealli, F., Pudney, S. and Thomas J. M. (1996), *Training Duration and Post.Training Outcomes: A Duration-Limited Competing Risks Model*, The Economic Journal, Vol. 106, No 435, pp. 422-433.

Narendranathan, W. and Stewart, M. B. (1993), *Modelling the Probability of Leaving Unemployment: Competing Risks Models with Flexible Base-Line Hazards*, Applied Statistics, Vol. 42, No 1, pp. 63-83.

Nickell, S. (1979), *Estimating the probability of leaving unemployment*, Econometrica, Vol. 47, pp. 1249-1266.

Nickell, S. (1997), *Unemployment and labor Market Rigidities: Europe versus North America*, The Journal of Economic Perspectives, Vol. 11, No 3, pp. 55-74.

OECD (2002a), *Employment Outlook 2002*, Paris.

OECD (may 2004), *Female labour force participation: past trends and main determinants in OECD countries*, OECD Economics Department.

Paggiaro, A. and Trivellato, U. (2002), *Assessing the Effects of the Mobility Lists Programme by Flexible Duration Models*, in *Labour*, Vol. 16, No 2, pp. 235-266.

Payne, C. and Payne, J. (2000), *Early Identification of the Long-Term Unemployed*, PSI Research Discussion Paper No 4, London, PSI.

Ricciardi, L. (1991), *La disoccupazione di lunga durata in Italia: un'analisi dell'evidenza empirica nel periodo 1977-1989*, in *Economia & Lavoro*, Vol. 25, No 2, pp. 69-94.

Salant, S.W. (1977), *Search Theory and Duration Data: A Theory of Sorts*, The Quarterly Journal of Economics, Vol. 91, No 1, pp. 39-57.

Thomas, J.M. (1996), *On the interpretation of covariate estimates in independent competing-risks models*, Bulletin of Economic Research, Vol. 48, No 1, pp. 27-39.

Trivedi, P. K. and Alexander, J. N. (1989), *Reemployment Probability and Multiple Unemployment Spells: A Partial-Likelihood Approach*, Journal of Business & Economic Statistics, Vol. 7, No 3, pp. 395-401.

Wooldridge, J.M. (2002), Econometric Analysis of Cross Section and Panel Data, Cambridge, MA, MIT Press.

Appendix

Table A1: Quarterly Unemployment rates by sex, duration, age groups and education (2004-2005), Lombardy's labour market.

	2004				**2005**			
Variable	**I quarter**	**II quarter**	**III quarter**	**IV quarter**	**I quarter**	**II quarter**	**III quarter**	**IV quarter**
				Males and Females				
Total	4.1	3.7	4.1	4.2	3.8	4.0	3.9	4.5
LTU[a]	1.4	1.4	1.6	1.6	1.3	1.4	1.2	1.8
Aged 15-24	11.7	11.9	13.9	13.5	10.2	11.6	13.6	14.6
Aged 25+	3.4	2.9	3.2	3.4	3.3	3.4	3.1	3.5
Education								
Not educ[b]	4.3	18.4	12.1	10.7	9.2	10.6	11.1	10.7
Less educ[c]	6.8	3.7	4.4	4.2	3.4	4.0	3.7	5.0
Primary	4.7	4	4.3	5.4	5.1	4.9	4.9	5.3
Secondary	3.8	3.7	3.9	3.7	2.9	0.7	3.5	3.9
Degree1	0.5	3.6	3	0.8	0.4	3.2	0.4	1.9
Degree2	2.3	2.5	0.9	2.1	2.1	2.4	3.4	3.0
PhD	1.3	0	1.8	2.0	0	0	0.4	0
				Males				
Total	3.2	2.6	2.6	3.1	2.8	3.2	3.1	3.5
LTU*	1	0.9	0.8	0.9	0.9	0.9	0.8	1.1
Aged 15-24	11.8	10.7	11.5	13.1	8.9	10.6	9.2	10.8
Aged 25+	2.5	1.9	1.8	2.2	2.3	2.5	2.6	2.7
Education								
Not educ*	2.7	13.0	5.6	5.6	1.3	3.9	5.5	4.5
Less educ*	5.2	2.7	3.5	3.6	2.9	3.4	2.3	3.7
Primary	3.1	2.7	2.4	3.8	3.4	3.8	3.8	4.2
Secondary	3.3	2.6	3.1	2.7	2.4	2.7	2.4	3.5
Degree1	2.0	0	2.6	0	0	1.0	0	0.5
Degree2	2.1	2.6	2.4	1.8	0.9	2.0	3.3	2.3
PhD	2.5	0	0	0.5	0	0	0.5	0
				Females				
Total	5.2	5.3	6.2	5.8	5.3	5.3	5.2	5.8
LTU*	1.9	2.1	2.7	2.6	1.9	2.1	1.8	2.7
Aged 15-24	11.6	13.4	16.6	13.9	11.9	12.8	14.1	14.9
Aged 25+	4.6	4.4	5.1	4.9	4.7	4.5	3.9	4.5

Education								
Not educ*	11.5	11.0	13.9	14.2	15.0	15.8	15.3	13.0
Less educ*	9.7	1.2	5.9	5.2	4.1	5.2	6.5	7.7
Primary	7.4	6.6	7.6	8.4	8.1	7.0	6.9	7.9
Secondary	4.1	5.0	5.8	5.1	3.5	3.6	4.9	4.4
Degree1	0	5.4	5.6	1.3	0.5	4.4	3.2	3.7
Degree2	2.6	2.5	3.8	2.4	3.6	2.9	3.7	3.8
PhD	0	0	2.6	2.0	0	0	0.2	0

Notes: (i) All the unemployment rates are related to Lombardy's labour market.
(ii) The rates have been computed by using each quarterly dataset and the proper weights to obtain representative samples at a regional level.
(iii) [a] denotes the long term unemployment rate.
(iv) [b c] denote individuals with no education titles and with five years of schooling respectively.
(v) All results are obtained using STATA.

Table A2: Unemployment Duration: Estimated Parameters from Three Parametric Models (standard errors)

	Exponential			Cox PH			Log Logistic		
Var	**coeff.**	**z**	**P>|z|**	**coeff.**	**z**	**P>|z|**	**coeff.**	**z**	**P>|z|**
indiv	0.180 (0.165)	1.09	0.274	0.075 (0.169)	0.45	0.655	-0.092 (0.224)	-0.41	0.679
sex	0.495 (0.127)	3.89	0.000**	0.298 (0.127)	2.35	0.019*	-0.370 (0.160)	-2.31	0.021*
age	-0.046 (0.007)	-6.29	0.000**	-0.024 (0.007)	-3.16	0.002**	0.035 (0.009)	3.74	0.000**
mstatus	-0.768 (0.211)	-3.64	0.000**	-0.551 (0.212)	-2.60	0.009**	0.685 (0.280)	2.45	0.014*
child	-0.235 (0.158)	-1.48	0.139	-0.191 (0.209)	-1.18	0.236	0.258 (0.209)	1.24	0.216
URate	-0.147 (0.278)	-0.53	0.596	0.199	0.72	0.472	-0.352 (0.350)	-1.01	0.314
married	0.847 (0.247)	3.43	0.001**	0.489	2.00	0.046*	-0.667 (0.327)	-2.04	0.041*
north	-0.769 (0.220)	-3.50	0.000**	-0.515	-2.35	0.019*	0.568 (0.271)	2.09	0.036*
south	-0.715 (0.222)	-3.21	0.001**	-0.342	-1.53	0.125	0.320 (0.278)	1.15	0.250
west	-0.861 (0.209)	-4.12	0.000**	-0.527	-2.52	0.012*	-0.558 (0.257)	2.17	0.030*
educ2	0.631 (0.117)	5.37	0.000**	0.389	3.28	0.001*	-0.466 (0.152)	-3.06	0.002**
_cons	-1.419 (1.218)	-1.16	0.244	_	_	_	3.216	2.10	0.035*
Obs		823			823			823	
LogL		-838.02			-1757.2			-791.57	

Notes: (i) Standard errors in brackets.
(ii) * ** denote significance at 5% and 1% respectively.
(iii) The Cox PH model has no intercept (_cons).
(iv) All results are obtained using STATA.

Table A3: Competing Risks: Estimated Parameters from Three Parametric Models (standard errors)

	Exponential				Cox PH				Log Logistic			
Var	**Risk 1**	**Risk 2**	**Risk 3**	**Risk 4**	**Risk 1**	**Risk 2**	**Risk 3**	**Risk 4**	**Risk 1**	**Risk 2**	**Risk 3**	**Risk 4**
indiv	0.456	0.256	1.805	-0.336	0.319	0.123	0.075	-0.070	-0.562	-0.287	-0.092	0.478
	(0.197)*	(0.233)	(0.165)	(0.181)	(0.204)	(0.241)	(0.169)	(0.184)	(0.359)	(0.433)	(0.224)	(0.260)
sex	0.926	0.924	0.495	-0.161	0.655	0.618	0.298	-0.299	-1.243	-1.267	-0.370	0.426
	(0.148)**	(0.172)**	(0.127)**	(0.147)	(0.146)**	(0.169)**	(0.127)*	(0.147)*	(0.254)**	(0.302)**	(0.160)*	(0.201)*
age	-0.069	0.061	-0.046	-0.028	-0.352	-0.027	-0.023	-0.011	0.070	0.061	0.035	0.714
	(0.009)**	(0.011)**	(0.007)**	(0.007)**	(0.009)**	(0.010)**	(0.007)**	(0.007)	(0.015)**	(0.018)**	(0.009)**	(0.010)
mstatus	-0.520	-0.557	-0.768	0.108	-0.153	-0.186	-0.551	0.312	0.294	0.359	0.685	-0.529
	(0.287)	(0.332)	(0.211)**	(0.255)	(0.289)	(0.335)	(0.212)**	(0.258)	(0.486)	(0.581)	(0.280)*	(0.359)
child	-0.120	0.038	-0.235	-0.053	0.299	0.143	-0.191	-0.008	0.026	-0.172	0.258	0.114
	(0.196)	(0.237)	(0.159)	(0.171)	(0.202)	(0.243)	(0.161)	(0.171)	(0.350)	(0.428)	(0.209)	(0.244)
URate	-0.780	-0.293	-0.147	-0.952	-0.248	0.211	0.199	-0.632	0.350	-0.452	-0.352	0.928
	(0.330)**	(0.375)	(0.278)	(0.295)**	(0.323)	(0.365)	(0.277)	(0.294)*	(0.555)	(0.652)	(0.349)	(0.418)*
married	0.696	0.677	0.847	0.339	0.216	0.195	0.489	-0.214	-0.377	-0.359	-0.667	0.470
	(0.331)**	(0.384)	(0.247)**	(0.290)	(0.330)	(0.382)	(0.244)*	(0.292)	(0.559)	(0.666)	(0.327)*	(0.407)
north	-0.449	-0.739	-0.769	-0.375	-0.244	-0.268	-0.515	-0.104	0.216	0.664	0.568	0.069
	(0.256)	(0.272)**	(0.220)**	(0.258)	(0.255)	(0.271)	(0.271)*	(0.257)	(0.441)	(0.492)	(0.271)*	(0.344)
south	-1.009	-1.125	-0.715	-0.517	-0.466	-0.567	-0.342	-0.184	0.994	1.182	0.319	0.226
	(0.284)**	(0.299)**	(0.222)**	(0.265)*	(0.282)	(0.298)*	(0.223)	(0.265)	(0.482)*	(0.534)*	(0.278)	(0.358)
west	-0.669	-0.979	-0.861	-0.609	-0.124	-0.374	-0.527	-0.302	0.457	0.942	0.558	0.468
	(0.247)**	(0.261)**	(0.209)**	(0.252)*	(0.247)	(0.260)	(0.209)*	(0.252)	(0.425)	(0.472)*	(0.257)*	(0.336)
educ2	0.692	0.566	0.631	0.078	0.309	0.184	0.389	-0.090	-0.647	-0.453	-0.465	0.186
	(0.137)**	(0.157)**	(0.117)**	(0.121)	(0.138)	(0.157)	(0.118)**	(0.122)	(0.241)	(0.282)	(0.152)**	(0.175)
_cons	-1.105	-1.224	-1.419	1.291	_	_	_	_	0.828	4.552	3.216	-1.415
	(1.457)	(1.661)	(1.218)	(1.288)					(2.399)	(2.841)	(1.528)	(1.786)
Obs	823	823	823	823	823	823	823	823	823	823	823	823
LogL	-849.28	-707.85	-838.02	-802.58	-1372.23	-1048.77	-1757.23	-1570.39	-770.93	-640.73	-791.58	-768.90

Notes: (i) Standard errors in brackets.
(ii) * ** denote significance at 5% and 1% respectively.
(iii) The Cox PH model has no intercept (_cons).
(iv) All results are obtained using STATA.