

\\58\\

**Modelling Employment Spells from Emilian
Labour Force Data**

by

Michele Lalla

June 1990

Dipartimento di Economia Politica
Via Giardini 454
41100 Modena (Italy)

Modelling Employment Spells from Emilian Labour Force Data

Michele LALLA, *Università di Modena*
Dipartimento di Economia Politica
Via P. Giardini n. 454 (Direz. 70), 41100 Modena, Italia

Abstract

The analysis of employment spells obtained from a Cross Population Survey (CPS) should take into account the problem of length-bias because CPS-type data are generally right-censored. In this paper, we examine the procedure to analyse a data set which includes only censored or incomplete spells. First of all, the distributions of the incomplete spells of employment are analysed without explanatory variables (such as age, sex, etc.), assuming that the unobserved completed spells have a Weibull, a log-logistic, or a Gompertz distribution, and the best assumption is adopted (the Weibull). Using the constant and time-dependent proportional hazards models, the methodology to study the effects of explanatory variables on spell distributions is then described for both incomplete spells and incomplete spells considered as completed ones. These methods are then applied to the lengths of ongoing spells of employment for employed and self-employed workers.

Keywords

Incomplete spells, completed spells, hazard function, employment exit rate, proportional hazards, constant covariates, time-dependent covariates.

I. Introduction¹

The duration data sets used to study employment or unemployment often come from samples of individuals at a given point in time (Current Population Survey) and they contain the lengths of ongoing spells of employment or unemployment for sample subjects who are currently employed or unemployed. These observations are right-censored and are called *spells in progress* or *incomplete spells*. They are defined as the length of time between the moment an individual finds a job and the survey date. The measured durations are generally retrospective data which involve the *telescoping effect*. Moreover, the sample may

¹ The author is indebted to Sebastiano Brusco, Andrea Gavioli, Nicola Torelli, and Giovanni Solinas for their helpful comments on a preliminary version of this paper. The usual *caveat* holds: the author is responsible for any errors. The data were processed at the CICAIA of the University of Modena.

be affected by an overrepresentation of long or short spells as the probability of their being selected is different and varies according to the shapes of their distribution (*length-biased sampling*). Moreover, there are difficulties in using the duration of employment as a dependent variable in a regression context where the explanatory variables change during employment spells. These drawbacks make it difficult to analyse duration data by traditional linear or nonlinear models.

Duration data may be analyzed fairly well using the survival statistical methods which allow one to process a set of data containing only completed spells, or both completed and incomplete spells, or only incomplete spells. Furthermore, the models constructed involve a stochastic process, which seems close to reality. In these methods, the hazard function, $h(t)$, plays the leading role, emphasizing the conditional probabilities². The unconditional probability, $f(t)$, of an event taking place may also be used because it involves the same parameters. Both conditional and unconditional probabilities are just two different ways of describing the same process. However, the hazard function, which represents the phenomenon in terms of a sequence of conditional probabilities, is quite appealing because it enables one to consider duration dependence immediately (state dependence) and/or heterogeneity (Heckmann, 1981; Flinn and Heckmann, 1982).

In this paper, we present the methods to analyse a data set of incomplete spells assuming a known distribution for the completed spells, but this assumption involves some difficulties, particularly in modelling the length of employment. First, population changes over time may modify the distribution of the spells, so studying only cohorts which contain homogeneous individuals is suggested. On the other hand it does not solve the problem because the cohort itself undergoes a change over time. Second, the stoppage of job relations when individuals reach retirement age, which is the same for all employed workers, forcibly interrupts the spells in progress, so one might suspect that probabilistic models become inadequate. Retirement is relevant when the parameters of the distribution function are estimated without explanatory variables (such as age, sex, sector, *etc.*). However, these aspects are inherent in all duration processes and in all distributions.

The description of the models adopted and the empirical results reported here concern the processing of incomplete spells only. First of all, an explorative analysis is described in which completed but unobserved spells are assumed to have a Weibull, a log-logistic, or a Gompertz distribution³ and the best assump-

² Although use of the life table method to analyse labour turnover dates back to Silcock (1954), this approach was explicitly suggested by Nickell (1979a, b) and Lancaster (1979) to study unemployment durations. Since then, a great number of applications of these techniques to economic duration data have been produced (Lancaster and Nickell, 1980; Solon, 1985; Narendranathan, Nickell, and Stern, 1985; Flinn, 1986; Ham and Rea, 1987; Kiefer, 1988; Torelli and Trivellato, 1988; among others).

³ Other candidates, such as the gamma, logistic, and Makeham, did not produce satisfactory results.

tion is selected. The effects of the determinants of employment length on spell distribution are then examined, assuming that completed but unobserved spells have the previously selected distribution (the Weibull).

In Section II, we report the characteristics of the sample and the rule used to measure the length of employment spells. In Section III, assuming that unobserved completed spells have a Weibull, a log-logistic, or a Gompertz distribution, we describe the method used to estimate the parameters of the spell distribution and the resulting estimates for the employed and self-employed workers separately. In Section IV, assuming that unobserved completed spells are Weibull-distributed, we deal with models that include the explanatory variables for the employed and self-employed workers separately. Section V concludes the discussion with some comments and remarks.

II. The observed sample

The sample of almost 12,500 individuals from 4,677 households in the Italian Region of Emilia-Romagna was selected according to the same technique used by ISTAT, Italy's primary national data collection agency, with the individuals being interviewed in January, 1984. Personal characteristics and family composition were surveyed starting from January, 1983. The change in states over 52 weeks in 1983 relative to each individual were reconstructed (Brusco and Solinas, 1986). Hence, it was possible to obtain completed unemployment and employment spells, for which the beginning and the end of the spells are retrospective data. However, the durations analysed in this paper refer only to the spells in progress of individuals who were employed from January 1983 to January 1984 without changing jobs.

We distinguished between employed and self-employed workers because the labour laws and regulations for the two categories are quite different. The ongoing spells of the employed workers are measured by the time spent in the same firm (number of years), regardless of any changes in duties and/or levels. The ongoing spells of the self-employed workers are measured by the time spent carrying out the same economic activity, regardless of place.

Those workers who had contracts with a fixed date of expiration were left out of the sample of employed workers because the end of their spells does not occur at random over time and they could introduce some bias in the estimates of the means of the durations. In fact, we are investigating long-term employment to learn something about the importance of labour relations in the work history of individuals. It is thus necessary to examine the core of the labour force.

III. The density function of the incomplete spells

Assuming that the duration of employment is a positive random variable,

the probability that a worker will change states is given by

$$\frac{S'(t)}{S(t)} = \frac{f(t)}{1 - F(t)} = -h(t) \quad (1)$$

where $f(t)$ is the probability density, $S(t)$ is the survivor function, $F(t)$ is the distribution function, and $h(t)$ is the hazard function. It is worth noting that $h(t)$ is the conditional instantaneous rate and hereafter it will be referred to as the employment exit rate. The solution of the previous equation is

$$F(t) = 1 - e^{-\int_0^t h(u)du} = 1 - e^{-H(t)} \quad (2)$$

where $H(t)$ denotes the integrated hazard. Only completed spells may be processed by this model. Data sets that also include incomplete spells may still be processed with some adjustments (Aitkin and Clayton, 1980; Cox and Oakes, 1984). If only spells in progress make up the data set, the distribution function of the incomplete spells $G(t)$ or the relative density $g(t)$ must be found. Let $f(s)$ be the density of the completed spells that begin at any time and let $g(s|t)$ be the density of the completed spells conditional on their being in progress at the time of the survey. Then, the following holds (Salant, 1977):

$$g(s|t) = \frac{sf(s)}{E(s)} \quad (3)$$

where $g(s|t)$ expresses the probability that completed spells of length s will be captured by the survey. Assuming that the beginning of the spell is equally likely to occur at any time and that the spell is drawn from the same distribution regardless of its starting date, the probability of the survey taking place at a given point in a spell of length s is $1/s$. If the employment spell distribution is truly stable and the changes in the rate of employment reflect only its incidence, we may have some spell length-bias, but it is not too serious a problem in practise. Some experiments with changes in the parameters over time have shown that the problem can be overlooked to some extent. Hence, the probability of a spell of length s being surveyed at a point t units from the beginning of s is $(1/s) \cdot [sf(s)/E(s)]$, $s \geq t$ (Salant, 1977). To calculate the probability that a spell captured in a survey is an incomplete spell of length t , we must add up all the favorable events:

$$g(t) = \int_t^\infty \frac{1}{s} \left[\frac{sf(s)}{E(s)} \right] ds = \frac{1}{E(s)} \int_t^\infty f(s) ds. \quad (4)$$

III.a. The examined density of the incomplete spells

The probability density of the incomplete spells (4) is analytic and easy to handle only when the distribution of completed spells is in turn analytic. The

functions used to fit the distribution of the employment spells are reported in Table 1.

Table 1 – *The cumulative distribution function, the probability density function, the hazard function, the mean, and the incomplete spell probability density function for the Weibull, the log-logistic, and the Gompertz.*

	Weibull	Log-logistic	Gompertz
$F(s)$	$1 - \exp[-(\lambda s)^\alpha]$	$1/[1 + (\lambda s)^\alpha]$	$1 - \exp\{-\lambda \exp[\alpha s]\}$
$f(s)$	$\lambda \alpha (\lambda s)^{\alpha-1} \exp[-(\lambda s)^\alpha]$	$[\lambda \alpha (\lambda s)^{\alpha-1}] / \{[1 + (\lambda s)^\alpha]^2\}$	$\lambda \alpha \exp[\alpha s] \cdot \exp\{-\lambda \exp[\alpha s]\}$
$h(s)$	$\alpha \lambda (\lambda s)^{\alpha-1}$	$[\lambda \alpha (\lambda s)^{\alpha-1}] / [1 + (\lambda s)^\alpha]$	$\lambda \alpha \exp[\alpha s]$
μ_S	$\Gamma(1/\alpha) / (\lambda \alpha)$	$\Gamma(1 - 1/\alpha) \Gamma(1 + 1/\alpha) / \lambda$	$[\Gamma'(1) - \log(\lambda)] / \alpha$
$g(t)$	$[(\lambda \alpha) / \Gamma(1/\alpha)] \exp[-(\lambda t)^\alpha]$	$\lambda / \{\Gamma(1 + 1/\alpha) \Gamma(1 - 1/\alpha) \cdot [1 + (\lambda t)^\alpha]\}$	$\{\alpha / [\Gamma'(1) - \log(\lambda)]\} \cdot \exp\{-\lambda \exp[\alpha t]\}$

The Weibull distribution, which depends on a shape ($\alpha > 0$) and a scale ($\lambda > 0$) parameter, allows for immediate interpretation of the hazard. If $\alpha > 1$, the employment exit rate is increasing with duration and the density tends to concentrate around the characteristic life ($1/\lambda$). If $\alpha < 1$, the employment exit rate is decreasing with duration. If $\alpha = 1$, the employment exit rate is constant, implying that the past is uninfluential on the future (exponential distribution). Thus, the Weibull is a generalization of the exponential distribution.

The probability density of the incomplete spells, $g(t)$, is expressed fairly only when the completed spells are Weibull-distributed, and it yields a mean $\mu_T = [\lambda \Gamma(1/\alpha)]^{-1} \Gamma(2/\alpha)$ and a variance $\sigma_T^2 = [\lambda \Gamma(1/\alpha)]^{-2} [\Gamma(3/\alpha) \Gamma(1/\alpha) - \Gamma^2(2/\alpha)]$ in an analytic form. Considering the mean of the Weibull distribution, reported in the Table 1, the following relation holds between the mean of the incomplete spells, μ_T , and the mean of the completed ones, μ_S ,

$$\frac{\mu_T}{\mu_S} = \frac{\lambda_S \alpha_S}{\lambda_T} \frac{\Gamma(\frac{2}{\alpha_T})}{\Gamma(\frac{1}{\alpha_T}) \Gamma(\frac{1}{\alpha_S})}. \quad (5)$$

As the parameters of the density of the full lengths of spells are the same as those for the density of the lengths of spells in progress, we obtain:

$$\frac{\mu_T}{\mu_S} = \alpha \frac{\Gamma(\frac{2}{\alpha})}{\Gamma^2(\frac{1}{\alpha})}. \quad (6)$$

Then, $\alpha > 1 \Rightarrow \mu_T < \mu_S$ and $\mu_T/\mu_S \rightarrow 1/2$ as $\alpha \rightarrow \infty$; $\alpha < 1 \Rightarrow \mu_T > \mu_S$ and $\mu_T/\mu_S \rightarrow \infty$ as $\alpha \rightarrow 0$; $\alpha = 1 \Rightarrow \mu_T = \mu_S$ (see Figure 1)⁴.

⁴ The methodology described above can also be used to analyse unemployment spells

Figure 1 - The ratio between the means of the incomplete and completed spells of employment according to the Weibull

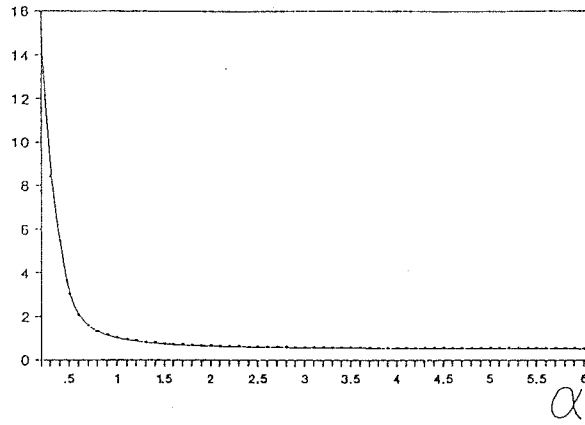


Figure 2 - The ratio between the means of the incomplete and completed spells of employment according to the log-logistic

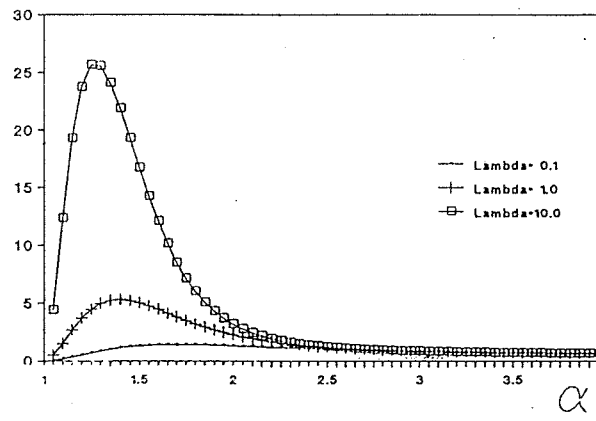
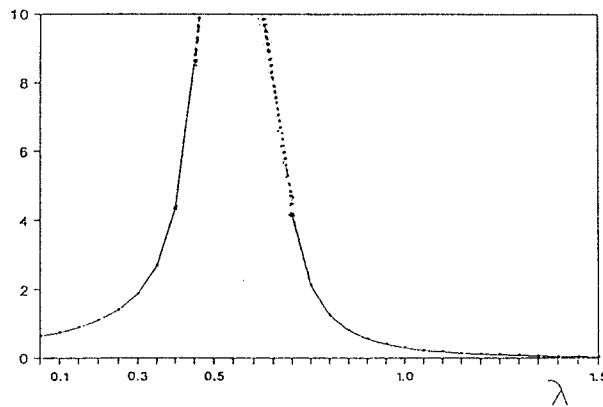


Figure 3 - The ratio between the means of the incomplete and completed spells of employment according to the Gompertz



in progress. In practise, the latter are often transformed into completed spells by doubling the lengths. The results obtained show that the mean of the lengths of completed spells is not equal to twice the mean of the lengths of the spells in progress. Salant (1977) provided for its general demonstration.

The expected length of the incomplete spells can only be calculated numerically when the completed ones have a log-logistic or a Gompertz distribution. The ratio μ_T/μ_S relative to the log-logistic, plotted as a function of α , showed different patterns with different values of the location parameter, λ (see Figure 2). The ratio μ_T/μ_S , calculated for the Gompertz and plotted as a function of λ , yielded the same curve for α varying between 0.01 and 0.35; a maximum ($\mu_T/\mu_S \simeq 1000$) was attained near $\lambda = 0.55$ (see Figure 3).

III.b. The method of estimation

The estimates are obtained by maximization of the log-likelihood, L , for a set of n independent observations: t_1, \dots, t_n . If the completed spells have a Weibull, a log-logistic, or a Gompertz distribution, then the log-likelihoods of the density function, $g(t)$, of the incomplete spells are respectively:

$$L = n \log(\lambda) - n \log[\Gamma(1 + 1/\alpha)] - \sum_{i=1}^n (\lambda t_i)^\alpha, \quad (7)$$

$$L = n \log(\lambda) - n \log[\Gamma(1 + 1/\alpha)] - n \log[\Gamma(1 - 1/\alpha)] - \sum_{i=1}^n \log[1 + (\lambda t_i)^\alpha], \quad (8)$$

$$L = n \log(\alpha) - n \log[\Gamma'(1) - \log(\lambda)] - \lambda \sum_{i=1}^n \exp(-\alpha t_i). \quad (9)$$

The estimates, which will be termed unbiased because the densities take right-censoring into account, emerge as a solution of the (nonlinear) system of the partial differential equations of the first order in λ and α . The solutions were obtained by NAG library routines. Variances of the estimates were determined by the expected value of the diagonal elements of the inverse of the Fisher information matrix which is defined by the second partial derivatives with signs changed and calculated at the point defined by the maximum likelihood estimates (Rao, 1973).

The maximum likelihood method is again adopted to estimate the parameters of the distribution of employment durations in the case where the spells in progress are assumed to be completed and the estimates will be termed biased. The log-likelihood functions of completed spells distributed as a Weibull, a log-logistic, and a Gompertz are respectively:

$$L = n \alpha \log(\lambda) + n \log(\alpha) + \sum_{i=1}^n [(\alpha - 1) \log(s_i) - (\lambda s_i)^\alpha], \quad (7')$$

$$L = n \alpha \log(\lambda) + n \log(\alpha) + \sum_{i=1}^n \{(\alpha - 1) \log(s_i) - 2 \log[1 + (\lambda s_i)^\alpha]\}, \quad (8')$$

$$L = n \log(\lambda) + n \log(\alpha) + \sum_{i=1}^n [\alpha s_i - \lambda \exp(\alpha s_i)]. \quad (9')$$

III.c. The results

The unbiased and biased estimates, the corresponding standard deviations, the log-likelihood, and the proportion of the total variation explained⁵ for each model estimated by the methods previously described are reported in Table 2. The best results were provided by the Weibull. Therefore, the comments are restricted to the latter.

The unbiased estimates differed significantly from the biased ones, implying that the mean of the employment spell durations estimated from the incomplete spells was greater than the mean of the incomplete spells considered as completed ones. In fact, the mean of the incomplete spells was $\hat{\mu}_T = [\hat{\lambda}\Gamma(1/\hat{\alpha})]^{-1}\Gamma(2/\hat{\alpha}) = 10.6$ years for the employed workers and $\hat{\mu}_T = 15.5$ years for the self-employed workers. The variances were $\hat{\sigma}_T^2 = [\hat{\lambda}\Gamma(1/\hat{\alpha})]^{-2}[\Gamma(3/\hat{\alpha})\Gamma(1/\hat{\alpha}) - \Gamma^2(2/\hat{\alpha})] = 65.4$ and $\hat{\sigma}_T^2 = 138.0$, respectively. The mean of the completed spells was then calculated by equation (6) or by the expression relative to the distribution assumed. The values, $\hat{\mu}_S = 16.3$ for employed workers and $\hat{\mu}_S = 24.2$ for self-employed workers, indicate an explicit job "tenure" in Italy and some rigidities in the Italian labour market, perhaps because a number of laws and regulations limit the possibility of dismissing workers.

The estimates of the parameters of the Weibull distribution of the employment spells obtained for Emilia-Romagna are different from those obtained by Flinn (1986) for Italy using ISTAT data collected in April, 1983. It is worth noting that the difference between the shape parameters considerably diminishes when an unobservable heterogeneity is introduced, although the means are not comparable because Flinn started from unemployment spells and used a model with highly unrealistic assumptions.

⁵ The proportion of the total variation explained is given by $R^2 = 1 - RSS/TSS$, where RSS indicates the sum of the squares of the differences between estimated and observed densities (residuals) and TSS indicates the sum of the squares of the differences between the observed densities and their mean (total).

Table 2 – Estimates of the Gompertz, Log-logistic, and Weibull Parameters on Right-Censored Durations of Employment, in Years, with and without Censoring Adjustment.

Density	n	Unbiased Estimates			Biased Estimates		
		$\hat{\lambda}_U$ (S. D.)	$\hat{\alpha}_U$ (S. D.)	L R^2	$\hat{\lambda}_B$ (S. D.)	$\hat{\alpha}_B$ (S. D.)	L R^2
Gompertz							
Employed workers	2995	3E-4 (1E-5)	0.297 (0.001)	-11064 0.620	0.212 (0.007)	0.104 (0.001)	-11137 0.338
Self-employed workers	1772	0.004 (3E-4)	0.149 (0.001)	-6797 0.480	0.205 (0.008)	0.073 (0.001)	-7234 0.141
Log-logistic							
Employed workers	2995	0.057 (0.001)	4.789 (0.085)	-9973 0.838	0.127 (0.002)	1.818 (0.035)	-10114 0.758
Self-employed workers	1772	0.038 (0.001)	5.011 (0.119)	-6577 0.585	0.088 (0.002)	1.705 (0.043)	-6553 0.484
Weibull							
Employed workers	2995	0.054 (0.001)	1.921 (0.080)	-9926 0.888	0.087 (0.001)	1.301 (0.019)	-9909 0.862
Self-employed workers	1772	0.037 (0.001)	1.976 (0.116)	-6554 0.636	0.060 (0.001)	1.257 (0.024)	-6554 0.606
Flinn : Italy without heterogeneity	1035	0.084 (0.003)	1.024 (0.051)	-8678			
Flinn : Italy (λ_{1S}) with heterogeneity (λ_{2S})	1035	0.027 (0.003)	1.868 (0.198)	-8664			
		0.081 (0.006)					

IV. The density of the incomplete spells with explanatory variables

The influences of the characteristics of the employed workers on the distribution of their spells can be formulated in terms of the effects of the covariates on the employment exit rate (Cox, 1972). This requires two assumptions:

- 1) the covariates affect the employment exit rate in a log-linear form;
 - 2) the employment exit rate is expressed by the underlying hazard function multiplied by the log-linear function of the covariates (proportionality).
- Then the hazard function, $h(t; \mathbf{X})$, for an individual with covariates \mathbf{X} , is

$$h(t; \mathbf{X}) = h_0(t) \exp(\mathbf{X}\beta) \quad (10)$$

where β is the vector of unknown parameters associated with covariates \mathbf{X} , and $h_0(t)$ is an unknown hazard for an individual with covariates $\mathbf{X} = 0$.

The coefficients β indicate the relationship between the covariates and the hazard function. A positive coefficient implies an increase in the hazard or employment exit rate and therefore a negative relationship with time. A negative coefficient implies the opposite effect: a decrease in the employment exit rate and a positive relationship with time.

Assuming that $h_0(t)$ is the hazard of the Weibull (see Table 1), equation (2) takes the form

$$F(t; \mathbf{X}) = 1 - \exp[-(\lambda t)^\alpha \exp(\mathbf{X}\beta)] \quad (11)$$

with a mean $\mu_S = [\lambda \alpha \exp(\mathbf{X}\beta/\alpha)]^{-1} \Gamma(1/\alpha)$ and variance $\sigma_S^2 = [\lambda \alpha \exp(\mathbf{X}\beta/\alpha)]^{-2} [2\alpha \Gamma(2\alpha) - \Gamma^2(1/\alpha)]$.

Assuming a Weibull distribution for the full lengths of spells, the density of the incomplete spells will be

$$g(t; \mathbf{X}) = \frac{\lambda}{\Gamma(1 + 1/\alpha)} \exp[\mathbf{X}\beta/\alpha - (\lambda t)^\alpha \exp(\mathbf{X}\beta)] \quad (12)$$

with a mean $\mu_T = [\lambda \Gamma(1/\alpha) \exp(\mathbf{X}\beta/\alpha)]^{-1} \Gamma(2/\alpha)$ and variance $\sigma_T^2 = [\lambda \Gamma(1/\alpha) \exp(\mathbf{X}\beta/\alpha)]^{-2} [\Gamma(3/\alpha) \Gamma(1/\alpha) - \Gamma^2(2/\alpha)]$. The method of maximum likelihood can be used again to estimate the parameters of $g(t; \mathbf{X})$ which are unbiased because this density takes right-censoring into account. The log-likelihood is:

$$L = n \log(\lambda) - n \log[\Gamma(1 + 1/\alpha)] + \sum_{i=1}^n [\mathbf{X}_i \beta / \alpha - (\lambda t_i)^\alpha \exp(\mathbf{X}_i \beta)]. \quad (13)$$

The maximum likelihood estimates, $\hat{\lambda}$, $\hat{\alpha}$, and $\hat{\beta}$, emerge as solutions of the likelihood equations, namely

$$\begin{aligned} \frac{\partial L}{\partial \lambda} &= n - \alpha \sum_{i=1}^n (\lambda t_i)^\alpha \exp(\mathbf{X}_i \beta) = 0 \\ \frac{\partial L}{\partial \alpha} &= \frac{n}{\alpha^2} \frac{\Gamma'(1 + 1/\alpha)}{\Gamma(1 + 1/\alpha)} - \frac{1}{\alpha^2} \sum_{i=1}^n \mathbf{X}_i \beta - \sum_{i=1}^n (\lambda t_i)^\alpha \exp(\mathbf{X}_i \beta) \log(\lambda t_i) = 0 \quad (14) \\ \frac{\partial L}{\partial \beta_j} &= \sum_{i=1}^n \frac{X_{ij}}{\alpha} - \sum_{i=1}^n (\lambda t_i)^\alpha \exp(\mathbf{X}_i \beta) X_{ij} = 0 \quad j = 1, \dots, k \end{aligned}$$

where k is the number of covariates. Variances of the estimates were obtained, again, by the expected value of the diagonal elements of the inverse of the Fisher information matrix.

When the spells in progress are considered as completed spells, the log-likelihood of the Weibull proportional hazards model which gives biased estimates, is

$$L = n \alpha \log(\lambda) + n \log(\alpha) + \sum_{i=1}^n [(\alpha - 1) \log(t_i) + \mathbf{X}_i \boldsymbol{\beta} - (\lambda t_i)^\alpha \exp(\mathbf{X}_i \boldsymbol{\beta})]. \quad (15)$$

IV.a. Applications: the case of employment spells

In modelling employment duration, it is difficult to take into account the forced interruption of an employment spell when the individual reaches retirement age. This situation may be relevant for employed workers because in compliance with labour laws, employers superannuate workers over 65 years of age; on the other hand, it may be irrelevant for self-employed workers because the labour laws permit them to work after the standard retirement age. In fact, the maximum age of employed workers was 68 years, whereas the maximum age of self-employed workers was 79 years. Minimum age for both was 15 years.

The probability of leaving a job was found to be empirically decreasing with duration (Flinn, 1986). This situation can be well represented by the model, but if the workers are 65 years old, we are almost certain that the majority will retire by the following year. Retirements presumably begin before this age; therefore two models were considered: one with constant and another with time-dependent explanatory variables. A polynomial, containing the terms age and squared age, was inserted in both models to account for age-related changes in hazard. Another polynomial, containing terms t and t^2 , was included in the time-dependent specification to provide functions of the entire previous history of the employment process.

IV.b. The duration of the spells of the employed workers

The model pertaining to employed workers includes age, gender, education level, job level, size of firm, and sector⁶. Table 3 includes the unbiased and biased estimates of the coefficients in the proportional hazards model with constant and time-dependent explanatory variables, as well as the least-squares estimates of the regression coefficients where the duration is the regredend.

⁶ Gender was introduced as a dummy (0 for males and 1 for females). The education level was an ordinal variable which was assigned 1 for illiterate or no schooling, 2 for elementary school education or low school level, 3 for middle school level, 4 for trade or vocational school (*diploma professionale*), 5 for high school (*diploma di maturità*), 6 for college education (*laurea*). Job grade was also expressed by dummies (unskilled and semiskilled workers, skilled workers, nonmanual workers), as was the economic sector (agriculture, manufacturing, and services).

The evaluation of the model's goodness of fit is given by $R^2 = 1 - L_R^{2/n}$, where L_R indicates the likelihood-ratio between the restricted ($\beta = 0$) and the unrestricted ($\beta \neq 0$) maximum of the likelihood functions (Ricolfi, 1984). This equality is well demonstrated in the regression context, but with some caution it may also be used to provide an idea of the fit. For each model, L_R comes from the ratio between the exponential of the relative log-likelihood in Table 2 and the exponential of its log-likelihood.

In the unbiased proportional hazards model with constant covariates, the scale and the shape parameters were greater than their corresponding values reported in Table 2. Therefore, the conditional probability of remaining on the same job sharply decreases as employment duration increases when individual and job characteristics are included in the model. The explanatory variables modify this probability. Age, size of firm, and the agricultural sector proved to be negative coefficients, which imply a decrease in the employment exit rate. Age was positively correlated with employment duration. Thus, it may be argued that difficulties in finding a new job and in using one's own skills increase employment spell lengths. As expected, the squared age coefficient was positive, implying a reverse effect for elderly workers. Size of firm revealed a decrease in the hazard because the protection of trade-union action becomes more incisive and/or job conditions improve as size increases. The mean of durations in the agricultural sector was higher than those of other sectors. Gender, educational level, job grade, and services proved to be positive coefficients which imply an increase in the employment exit rate. Females, workers with high levels of education, skilled workers, and unskilled-semiskilled workers had a higher probability of leaving employment than males, workers with low levels of education, and nonmanual workers, respectively. In the biased proportional hazards model with constant covariates, the estimated parameters were lower than the unbiased ones. In the linear regression, the relationships between the covariates and employment durations were generally the same as those in the unbiased constant model.

In the unbiased proportional hazards model with time-dependent covariates the coefficients also showed a pattern analogous to the unbiased constant model coefficients⁷. More specifically, the effect of employment duration length was similar to the that of age. So, the hazard increases when workers have long spells of employment and it decreases when they have short ones. In this case, the estimated parameters yielded values that were generally higher than others, as did the log-likelihood and R^2 .

⁷ The estimates of the unbiased time-dependent models were obtained using the T.S.P. ML procedure (Hall, 1983) for both the employed and the self-employed workers, whereas the estimated parameters of all the other models were derived from NAG library routines. It is worth noting that the maximum was not easily achieved in the time-dependent models because the surfaces are generally uneven.

Table 3 – *Unbiased and biased estimated parameters of the proportional hazards models (with constant and time-dependent covariates) and linear estimates relative to employment duration for employed workers.*

Covariates	$\hat{\beta}_U$ (S.D.)	$\hat{\beta}_B$ (S.D.)	$\hat{\beta}_{ols}$ (S.E.)	$\hat{\beta}_U$ (S.D.)	$\hat{\beta}_B$ (S.D.)
λ	1.076 (0.127)	3.024 (0.443)	—	1.805 (0.001)	0.610 (0.029)
α	6.587 (0.389)	1.859 (0.028)	—	10.537 (0.002)	6.067 (0.094)
Females	0.077 (0.102)	0.056 (0.040)	-0.582 (0.250)	0.991 (0.002)	-0.052 (0.040)
Education level	0.305 (0.072)	0.043 (0.026)	0.132 (0.164)	0.460 (0.001)	-0.022 (0.026)
Skilled workers	0.518 (0.146)	0.306 (0.058)	-2.391 (0.362)	0.714 (0.001)	-0.010 (0.059)
Semiskilled and unskilled workers	0.230 (0.142)	0.096 (0.056)	-0.991 (0.353)	0.507 (0.001)	-0.134 (0.057)
Size of firm	-0.006 (0.043)	-0.035 (0.017)	0.483 (0.106)	-0.333 (0.001)	0.021 (0.017)
Agriculture	-0.096 (0.276)	-0.168 (0.108)	1.445 (0.672)	-2.700 (0.002)	-0.054 (0.107)
Services	0.494 (0.172)	0.067 (0.067)	0.708 (0.410)	-0.437 (0.001)	0.037 (0.066)
Age/10	-8.423 (0.537)	-2.710 (0.123)	4.166 (0.725)	-8.437 (0.001)	-0.620 (0.141)
Age ² /100	0.680 (0.051)	0.230 (0.015)	0.045 (0.093)	0.621 (0.001)	0.075 (0.018)
T/10	—	—	—	-4.262 (0.002)	-10.790 (0.173)
T ² /100	—	—	—	0.604 (0.002)	1.661 (0.029)
L	-8925.6	-8934.4	-9737.2	1.0E+6	-5486.1
R^2	0.482	0.478	0.404	0.999	0.934

In the biased proportional hazards model with time-dependent covariates, the estimates of the scale and shape parameters were similar to corresponding estimates obtained in the simple unbiased proportional hazards model, whereas the estimated coefficients of the covariates often had different signs and values

lower than the corresponding unbiased ones. The effects of age and length of employment, as well as the fit, were not very different from the model that takes censoring into consideration.

IV.c. The durations of the spells of the self-employed workers

The employment duration models for the self-employed workers are similar to the previous ones. The models again include age, gender, education level, sector, and a dummy, which distinguishes the coadiuvants from the others. The estimated parameters of the models are reported in Table 4, as for the employed workers.

In the unbiased (constant) proportional hazards model, the scale and shape parameters revealed the same patterns described for the employed workers. Negative coefficients were estimated for age, the agricultural and services sectors. The coefficients of the agricultural and services sectors were different from those estimated for employed workers. Gender and educational level proved to be positive coefficients. The difference attributable to gender was significantly more marked amongst the self-employed workers. In the biased (constant) proportional hazards model, the estimated parameters were, once again, lower than the corresponding unbiased estimates. In the linear regression, the relationships with covariates remained the same as in the two previous models.

The unbiased time-dependent model also showed that the coefficients of the age and the duration polynomials had different signs. Thus, the employment exit rate increases when the workers approach retirement age or have long spells of employment, but it decreases when they are far from the critical age or they have short spells of employment. The shape and the location parameters, some coefficients of the covariates, the log-likelihood, and R^2 , were even greater than those obtained from the unbiased constant model. In the biased time-dependent model, the estimated parameters were always lower than the relative unbiased ones. The pattern revealed by R^2 , the effects of age, and the length of employment was similar to that of the unbiased time-dependent model.

Table 4 – Unbiased and biased estimated parameters of the proportional hazards models (with constant and time-dependent covariates) and linear estimates relative to employment duration for self-employed workers.

Covariates	$\hat{\beta}_U$ (S.D.)	$\hat{\beta}_B$ (S.D.)	$\hat{\beta}_{\alpha_s}$ (S.E.)	$\hat{\beta}_U$ (S.D.)	$\hat{\beta}_B$ (S.D.)
λ	0.647 (0.085)	2.124 (0.357)	—	1.534 (0.006)	0.447 (0.032)
α	7.433 (0.639)	1.907 (0.037)	—	8.263 (0.015)	5.120 (0.103)
Females	0.595 (0.156)	0.262 (0.055)	-2.362 (0.431)	1.750 (0.002)	0.015 (0.055)
Education level	0.283 (0.096)	0.078 (0.030)	-0.358 (0.244)	0.591 (0.001)	0.051 (0.030)
Coadiuvants	0.200 (0.274)	-0.030 (0.099)	0.040 (0.799)	-0.317 (0.002)	-0.031 (0.100)
Agriculture	-1.230 (0.192)	-0.580 (0.068)	6.637 (0.537)	1.164 (0.004)	0.117 (0.070)
Services	-0.071 (0.169)	-0.085 (0.059)	0.680 (0.471)	0.347 (0.003)	0.010 (0.059)
Age/10	-7.337 (0.619)	-2.324 (0.133)	2.509 (1.003)	-6.738 (0.006)	-0.691 (0.161)
Age ² /100	0.502 (0.049)	0.167 (0.014)	0.405 (0.113)	0.759 (0.001)	0.075 (0.018)
T/10	—	—	—	-7.409 (0.010)	-6.277 (0.131)
T ² /100	—	—	—	0.653 (0.001)	0.650 (0.015)
L	-5846.5	-5860.5	-6202.3	4.7E+5	-4138.5
R^2	0.543	0.543	0.543	0.999	0.935

V. Conclusions

The methodology described above is useful in the analysis of incomplete spells and is based upon the probability density that a spell captured in the survey has an incomplete length equal to the observed length, but it requires an *a priori* assumption referring to the distribution of completed spells which is unknown. However, if the assumed distribution does not have an analytical expression, the density of spells in progress will have terms too awkward to calculate.

We have treated two cases separately. The first concerns the procedure to estimate the parameters of the distribution function without the explanatory variables. The method described does not involve any terms too awkward to calculate and it may be easily applied to other distribution functions. The estimated parameters of the distribution with adjustment for the censoring, differed significantly from the estimated parameters of the distribution that does not consider the censoring. Therefore, it would be more appropriate to use the procedure which takes the spells in progress into consideration.

The second case concerns the analysis of the influences of the explanatory variables. The coefficients estimated by the constant proportional hazards model that takes right-censored into account, were different from those obtained without taking censoring into account: sometimes the signs were not equal and the values of the former were greater than the latter, thus emphasizing the impact of the covariates. The unbiased coefficients had signs similar to the linear estimates, though they were still biased. Therefore, if the aim is to verify the effect of the covariates on duration, the linear model is satisfactory and its application is straightforward.

The proportional hazards model with time-dependent covariates provided the best results, although it cannot be compared with the regression model. The computation of the estimates is much more difficult in this case. Nevertheless, empirical evidence suggests it would be convenient to introduce functions of the whole previous work history and evolutionary properties of the employment process.

References

- [1] M. Aitkin and D. Clayton, The Fitting of Exponential, Weibull and Extreme Value Distributions to Complex Censored Survival Data using GLIM, *Applied Statistics* **29** (1980) 156-163.
- [2] S. Brusco and G. Solinas, Mobilità occupazionale e disoccupazione in Emilia Romagna, *Materiali di discussione no. 17*. Dip. di Economia Politica, Università degli Studi, Modena, 1986.
- [3] D. R. Cox, Regression Models and Life Tables, *J. Roy. Statist. Soc.* **34 B** (1972) 187-220.
- [4] D. R. Cox and D. Oakes, *Analysis of Survival Data* (Chapman and Hall, London, 1984).
- [5] J. C. Flinn, Wages and Job Mobility of Young Workers, *J. P. E.* **94** (June 1986) S88-S110.
- [6] J. C. Flinn, Econometric Analysis of CPS-Type Unemployment Data, *J. H. R.* **XXI** (4) (1986) 456-484.
- [7] J. C. Flinn and J. J. Heckman, Models for the Analysis of Labour Force Dynamics, in: R. Basmann and G. Rhodes (Ed.), *Advances in Econometrics*, Vol. 1 (JAI, Greenwich, 1982).

- [8] B. H. Hall, *Reference Manual* (TSP International, Stanford, 1983b).
- [9] J. C. Ham and S. A. Rea Jr., Unemployment Insurance and Male Unemployment Duration in Canada, *J. Labour Economics* 5 (3) (1987) 325–353.
- [10] J. J. Heckman, Heterogeneity and State Dependence, in: Sherwin Rosen (Ed.), *Studies in Labor Markets*, (University of Chicago Press, Chicago, 1981).
- [11] N. M. Kiefer, Economic Duration Data and Hazard Functions, *J. Economic Literature* XXVI (June 1988) 646–679.
- [12] T. Lancaster, Econometric Methods for the Duration of Unemployment, *Econometrica* 47 (1979) 939–956.
- [13] T. Lancaster and S. Nickell, The Analysis of Re-Employment Probabilities for the Unemployed, *J. Roy. Statist. Soc.* 143 A (1980) 141–165.
- [14] W. Narendranathan, S. Nickell, and J. Stern, Unemployment Benefits Revisited, *Economic J.* 95 (June 1985) 307–329.
- [15] S. Nickell, Estimating the Probability of Leaving Unemployment, *Econometrica* 47 (1979a) 1249–1266.
- [16] S. Nickell, The Effect of Unemployment and Related Benefits on the Duration of Unemployment, *Economic J.* 89 (March 1979b) 34–49.
- [17] C. R. Rao, *Linear Statistical Inference and Its Applications* (Wiley, New York, 1973).
- [18] L. Ricolfi, *Modelli dell'attore e analisi dei dati* (Giappichelli, Torino, 1984).
- [19] S. W. Salant, Search Theory and Duration Data: A Theory of Sorts, *Q. J. E.* 91 (1) (1977) 39–57.
- [20] H. Silcock, The Phenomenon of Labour Turnover, *J. Roy. Statist. Soc.* CXVII A (1954) 429–440.
- [21] G. Solon, Work Incentive Effects of Taxing Unemployment Benefits, *Econometrica* 53 (2) (1985) 295–306.
- [22] N. Torelli and U. Trivellato, Modelling Job-Search Duration from the Italian Labour Force Data, *Labour* 2 (1) (1988) 117–134.

Mathematical symbols

Symbols	Description
α	shape parameter of the distribution functions
λ	scale parameter of the distribution functions
α_T	shape parameter of the incomplete spell distribution
λ_T	scale parameter of the incomplete spell distribution
α_S	shape parameter of the completed spell distribution
λ_S	scale parameter of the completed spell distribution
β	vectors of parameters associated with covariates
μ_T	mean of the incomplete spell distribution
μ_S	mean of the completed spell distribution
σ_T^2	variance of the incomplete spell distribution
σ_S^2	variance of the completed spell distribution
∞	infinite
\int_t^∞	integral from t to infinite
$\Gamma(\cdot)$	Gamma function
$\Gamma'(1)$	the first derivative of the Gamma function, $\Gamma(z)$, with respect to z at $z = 1$; $\Gamma'(1) = -0.5772157 \dots$

Materiali di discussione

1. Maria Cristina Marcuzzo [1985] "Joan Violet Robinson (1903-1983)", pp.134.
2. Sergio Lugaresi [1986] "Le imposte nelle teorie del sovrappiù", pp.26.
3. Massimo D'Angelillo e Leonardo Paggi [1986] "PCI e socialdemocrazie europee. Quale riformismo?", pp.158.
4. Gian Paolo Caselli e Gabriele Pastrello [1986] "Un suggerimento hobsoniano su terziario e occupazione: il caso degli Stati Uniti 1960/1983", pp.52.
5. Paolo Bosi e Paolo Silvestri [1986] "La distribuzione per aree disciplinari dei fondi destinati ai Dipartimenti, Istituti e Centri dell'Università di Modena: una proposta di riforma", pp.25.
6. Marco Lippi [1986] "Aggregation and Dynamics in One-Equation Econometric Models", pp.64.
7. Paolo Silvestri [1986] "Le tasse scolastiche e universitarie nella Legge Finanziaria 1986", pp.41.
8. Mario Forni [1986] "Storie familiari e storie di proprietà. Itinerari sociali nell'agricoltura italiana del dopoguerra", pp.165.
9. Sergio Paba [1986] "Gruppi strategici e concentrazione nell'industria europea degli elettrodomestici bianchi", pp.56.
10. Nerio Naldi [1986] "L'efficienza marginale del capitale nel breve periodo", pp.54.
11. Fernando Vianello [1986] "Labour Theory of Value", pp.31.
12. Piero Ganugi [1986] "Risparmio forzato e politica monetaria negli economisti italiani tra le due guerre", pp.40.
13. Maria Cristina Marcuzzo e Annalisa Rosselli [1986] "The Theory of the Gold Standard and Ricardo's Standard Commodity", pp.30.
14. Giovanni Solinas [1986] "Mercati del lavoro locali e carriere di lavoro giovanili", pp.66.
15. Giovanni Bonifati [1986] "Saggio dell'interesse e domanda effettiva. Osservazioni sul capitolo 17 della General Theory", pp.42.
16. Marina Murat [1986] "Between old and new classical macroeconomics: notes on Leijonhufvud's notion of full information equilibrium", pp.20.
17. Sebastiano Brusco e Giovanni Solinas [1986] "Mobilità occupazionale e disoccupazione in Emilia Romagna", pp.48.
18. Mario Forni [1986] "Aggregazione ed esogeneità", pp.13.
19. Sergio Lugaresi [1987] "Redistribuzione del reddito, consumi e occupazione", pp. 17.
20. Fiorenzo Sperotto [1987] "L'immagine neopopulista di *mercato debole* nel primo dibattito sovietico sulla pianificazione", pp. 34.

21. M. Cecilia Guerra [1987] "Benefici tributari del regime misto per i dividendi proposto dalla Commissione Sarcinelli: una nota critica", pp 9.
22. Leonardo Paggi [1987] "Contemporary Europe and Modern America: Theories of Modernity in Comparative Perspective", pp. 38.
23. Fernando Vianello [1987] "A Critique of Professor Goodwin's 'Critique of Sraffa' ", pp. 12.
24. Fernando Vianello [1987] "Effective Demand and the Rate of Profits: Some Thoughts on Marx, Kalecki and Sraffa", pp. 41.
25. Anna Maria Sala [1987] "Banche e territorio. Approccio ad un tema geografico-economico", pp. 40.
26. Enzo Mingione e Giovanni Mottura [1987] "Fattori di trasformazione e nuovi profili sociali nell'agricoltura italiana: qualche elemento di discussione", pp. 36.
27. Giovanna Procacci [1988] "The State and Social Control in Italy During the First World War", pp. 18.
28. Massimo Matteuzzi e Annamaria Simonazzi [1988] "Il debito pubblico", pp. 62.
29. Maria Cristina Marcuzzo (a cura di) [1988] "Richard F. Kahn. A disciple of Keynes", pp. 118.
30. Paolo Bosi [1988] "MICROMOD. Un modello dell'economia italiana per la didattica della politica fiscale", pp. 34.
31. Paolo Bosi [1988] "Indicatori della politica fiscale. Una rassegna e un confronto con l'aiuto di MICROMOD", pp. 25.
32. Giovanna Procacci [1988] "Protesta popolare e agitazioni operaie in Italia 1915-1918", pp. 45.
33. Margherita Russo [1988] "Distretto industriale e servizi. Uno studio dei trasporti nella produzione e nella vendita delle piastrelle", pp. 157.
34. Margherita Russo [1988] "The effects of technical change on skill requirements: an empirical analysis", pp. 28.
35. Carlo Grillenzoni [1988] "Identification, estimation of multivariate transfer functions", pp. 33.
36. Nerio Naldi [1988] "Keynes' concept of capital" pp. 40.
37. Andrea Ginzburg [1988] "Locomotiva Italia?" pp. 30.
38. Giovanni Mottura [1988] "La 'persistenza' secolare. Appunti su agricoltura contadina ed agricoltura familiare nelle società industriali" pp. 40.
39. Giovanni Mottura [1988] "L'anticamera dell'esodo. I contadini italiani dalla 'restaurazione contrattuale' fascista alla riforma fondiaria" pp. 40.
40. Leonardo Paggi [1988] "Americanismo e riformismo. La socialdemocrazia europea nell'economia mondiale aperta" pp. 120.
41. Annamaria Simonazzi [1988] "Fenomeni di isteresi nella spiegazione degli alti tassi di interesse reale" pp. 44.
42. Antonietta Bassetti [1989] "Analisi dell'andamento e della casualità della borsa valori" pp. 12.
43. Giovanna Procacci [1989] "State coercion and worker solidarity in Italy (1915-1818): the moral and political content of social unrest" pp. 41.
44. Carlo Alberto Magni [1989] "Reputazione e credibilità di una minaccia in un gioco bargaining"

pp. 56.

45. Giovanni Mottura [1989] "Agricoltura familiare e sistema agroalimentare in Italia" pp. 84.
46. Mario Forni [1989] "Trend, Cycle and 'Fortuitous Cancellations': a Note on a Paper by Nelson and Plosser" pp. 4.
47. Paolo Bosi, Roberto Golinelli, Anna Stagni [1989] "Le origini del debito pubblico e il costo della stabilizzazione" pp. 26.
48. Roberto Golinelli [1989] "Note sulla struttura e sull'impiego dei modelli macroeconomici" pp. 21.
49. Marco Lippi [1989] "A Short Note on Cointegration and Aggregation" pp. 11.
50. Gian Paolo Caselli and Gabriele Pastrello [1989] "The Linkage between Tertiary and Industrial Sector in the Italian Economy: 1951-1988. From an External Dependence to an Internal One" pp. 40
51. Gabriele Pastrello [1989] "François Quesnay: dal Tableau Zig-Zag al Tableau formule: una ricostruzione" pp. 48
52. Paolo Silvestri [1989] "Il bilancio dello stato" pp. 34
53. Tim Mason [1990] "Tre seminari di Storia Sociale Contemporanea" pp. 26
54. Michele Lalla [1990] "The Aggregate Escape Rate Analysed through the Queueing Model" pp. 23
55. Paolo Silvestri [1990] "Sull'autonomia finanziaria delle Università" pp. 11
56. Paola Bertolini, Enrico Giovannetti [1990] "Uno studio di 'filiera' nell'agroindustria. Il caso del Parmigiano Reggiano" pp. 164
57. Paolo Bosi, Roberto Golinelli, Anna Stagni [1990] "Effetti macroeconomici, settoriali e distributivi dell'armonizzazione dell'IVA" pp. 24