THE EGYPTIAN POPULATION AND
FAMILY PLANNING REVIEW.
ISSR,CAIRO UNIV. Vol (32),No 2,1999.

SURVIVAL ANALYSIS OF UNEMPLOYMENT WITH AN APPLICATION TO THE CASE OF EGYPT

BY

AMINA E. ABU-HUSSIEN

Department of Statistics, Faculty of Commerce,

AL-Azhar University, Girls Branch, Egypt

ABSTRACT

A statistical model is developed for describing the survival distribution of the waiting time to have a job for unemployed individuals. With such a model it is possible to make statistical inferences. It should be noted that in order to develop this model, one requires the lifetimes of all individuals under study. In many cases these lifetimes are not available. The model suggested in the present study introduces a method to overcome this difficulty. An application of the suggested model to the case of Egypt in 1996 is introduced.

Keywords and phrases: *survival analysis; gross years of working life; type I censored sample; constant hazard rate.*

1. INTRODUCTION

In any country the rate of unemployment is one of the most sensitive measures of social and economic development. Length of non-working life is a valuable measure in studies of unemployment. It represents the period of life spent outside the work force. It is affected mainly by the level and duration of labour force participation.

Unemployment is affected by many socio-economic variables; such as: educational level, marital status, sex, the presence of dependents in the household, the presence of other wage earners in the household, the number and type of jobs created for young people by economic development, school enrolment trends, the size of the old-age pension, and for women in general unemployment is affected by, age at marriage and number of children.

The econometric problems and methods involved in interpreting the causes of variation between unemployed job seekers in the duration of unemployment in the light of search theories have been studied by many economists [Lancaster (1979) and Balch and Bichaka (1997)].

Many studies investigate the effect of unemployment insurance on transition from unemployment to employment and on labour force with-drawl. The results of these studies suggest that unemployment insurance lengthens unemployment spells [Miles (1993), Poterla and Summers (1995) and Gourieroux and Scaillet (1997)].

The demographic literature is rich in studies concerned with unemployment. Most of these studies are descriptive and little attention is given to examine the duration of unemployment. [Addison and Portugal (1992), Andrew and Alan (1992), Fusun (1992), James (1995), Datta et. al. (1999) and Firth and Payne (1999)]. Working life tables are used in studying the loss of economically active life due to unemployment. [El-Biblawi (1990), El-Waffa`I. (1993)] .

The present study is concerned with the development of more efficient statistical methods for studying the duration of unemployment.

The aim of the present study is to develop what, we would call, a statistical model to represent the survival distribution of unemployment i.e. the distribution of the waiting time to have a job for unemployed individuals and to make statistical inferences on the basis of this model. It should be noted that in order to develop this model, one requires the lifetimes of all individuals under study. In many cases these lifetimes are not available. The model suggested in the present study introduces a method to overcome this difficulty and enables researchers to utilize such data. A theoretical framework is introduced in Section 2. The suggested model is presented in Section 3. In Section 4 an application of the suggested model to the case of Egypt in 1996 is introduced.

2. THEORETICAL FRAMEWORK

The times to the occurrences of events are termed " survival times " or " lifetimes " or " waiting times ". In the present study we are interested in the analysis of the survival times of unemployed individuals.

Let T be a continuous nonnegative random variable representing the survival times in the unemployment state for individuals in the population under study.

Let $f(t)$ denote the probability density function (p.d.f.) of T and let the distribution function be:

$$F(t) = P(T \leq t) = \int_0^t f(x)\,dx. \tag{2.1}$$

The probability of an individual surviving in the unemployment state till time t is called the survival function $S(t)$, where:

$$S(t) = P(T \geq t) = \int_t^{\infty} f(x)\,dx = 1 - F(t). \tag{2.2}$$

Note that $S(t)$ is a monotone decreasing continuous function with $S(0) = 1$ and $S(\infty) = \lim_{t \to \infty} S(t) = 0$.

The hazard function specifies the instantaneous probability of having a job at time t, given that the individual stayes in the unemployment state up till t, it is defined as

$$h(t) = \lim_{\Delta t \to 0} \frac{P\{t \leq T < t + \Delta t | T \geq t\}}{\Delta t} = \frac{f(t)}{S(t)} \tag{2.3}$$

i.e. the hazard function describes the way in which the instantaneous probability of having a job for an individual changes with time.

The hazard function $h(t)$ for a continuous survival time distribution possesses the following properties:

$$h(t) \geq 0 \quad \text{and} \quad \int_0^\infty h(t)dt = \infty .$$

Having determined $h(t)$, it is easy to derive expressions for $S(t)$ and $f(t)$ as follows:

$$S(t) = \exp\left\{-\int_0^t h(x)\,dx\right\}, \quad \text{and} \tag{2.4}$$

$$f(t) = h(t).\exp\left\{-\int_0^t h(x)\,dx\right\} . \tag{2.5}$$

All functions, unless stated otherwise, are defined over the interval $[0,\infty)$.

Unemployed individuals in a certain age group are considered a random sample censored at the end of this age group i.e. we have Type I censored sample .

Suppose that there are n individuals under study and:

T_i : is the survival time of the $i\underline{th}$ individual .

L_i : is a fixed censoring time for the $i\underline{th}$ individual .

The T_i's are assumed to be i.i.d. with p.d.f. $f(t)$ and survival function $S(t)$. The exact lifetime T_i of the $i\underline{th}$ individual will be observed only if $T_i \leq L_i$. The data in this case can be represented by the n pairs of random variables (t_i, γ_i) where $t_i = \min(T_i, L_i)$ and

$$\gamma_i = \begin{cases} 1 & \text{if } T_i \leq L_i \\ 0 & \text{if } T_i > L_i \end{cases} .$$

The joint p.d.f. of t_i and γ_i is $f(t_i)^{\gamma_i} S(L_i)^{1-\gamma_i}$.

The general form of the likelihood function if the pairs (t_i, γ_i) are independent for Type I censored data is

$$L = \prod_{i=1}^{n} f(t_i)^{\gamma_i} S(L_i)^{1-\gamma_i} . \tag{2.6}$$

For Type I censoring, exact small-sample test and interval estimation procedures are not easily obtainable, therefore methods based on large-sample properties of maximum likelihood are recommended [Bain (1978)]

In the case of constant hazard rate, the distribution of waiting time is exponential with the following p.d.f.

$$f(t) = \frac{1}{\theta} \exp\left\{-\frac{t}{\theta}\right\}, \qquad t \geq 0; \theta > 0, \tag{2.7}$$

for this exponential model, the likelihood function (2.6) in the case of Type I censored sample is

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} \left[\frac{1}{\theta} \exp\left\{ \frac{-T_i \gamma_i}{\theta} \right\} \right] \exp\left\{ -\frac{L_i(1-\gamma_i)}{\theta} \right\} \\ &= \frac{1}{\theta^r} \exp\left\{ -\sum_{i=1}^{n} \frac{t_i}{\theta} \right\}, \end{aligned} \tag{2.8}$$

where r is the observed number of lifetimes in the unemployment state, and the m.l.e. for θ is ;

$$\hat{\theta} = \frac{T}{r} \quad , \qquad r > 0 \tag{2.9}$$

where $T = \sum_{i=1}^{n} t_i$ is the total observed lifetime.

The Fisher information is :

$$I(\theta) = E\left[-\frac{d^2 \log L}{d\theta^2} \right] = \frac{1}{\theta^2} \sum_{i=1}^{n} \left[1 - \exp\left\{ -\frac{L_i}{\theta} \right\} \right]. \tag{2.10}$$

and the asymptotic normal approximation is :

$$\frac{\hat{\theta} - \theta}{I(\hat{\theta})^{-\frac{1}{2}}} \approx N(0,1). \tag{2.11}$$

is used to obtain confidence intervals or significance tests for θ . It should be noted that in order to calculate $I(\hat{\theta})$ one requires the censoring times

L_i of all individuals in the sample. In cases where these censoring times are not all available, the asymptotically equivalent approximation can be used :

$$\frac{\hat{\theta}-\theta}{I_0^{-\frac{1}{2}}} \approx N(0,1), \tag{2.12}$$

where

$$I_0 = \left[-\frac{d^2 \log L}{d\theta^2}\right]_{\theta=\hat{\theta}} = \frac{r}{\hat{\theta}^2} \tag{2.13}$$

is the observed information.

If $A(T) \leq \theta \leq B(T)$ is an α confidence interval for the mean survival time θ , then :

$$\frac{1}{B(T)} \leq \frac{1}{\theta} \leq \frac{1}{A(T)}$$

is an α confidence interval for the constant hazard rate , $\lambda = \frac{1}{\theta}$,

$$\exp\left\{-\frac{t_0}{A(T)}\right\} \leq S(t_0) \leq \exp\left\{-\frac{t_0}{B(T)}\right\}$$

is an α confidence interval for the survival function at $t_0, S(t_0)$,

$$[-\log(1-p)]A(T) \leq t_p \leq [-\log(1-p)]B(T)$$

is an α confidence interval for the $p\underline{th}$ quantile of the distribution , t_p, where:

$$t_p = \theta\, [-\log(1-p)].$$

3. THE SUGGESTED MODEL

3.1 Model Assumptions

1. Individuals in each age group are considered a random sample censored at the end of this age group i.e. we have Type I censored samples.

2. The case of a single lifetime variable, T, is considered, where T is a non-negative continuous random variable representing the lifetimes of individuals in the unemployment state in a homogeneous population.
3. The instantaneous probability of having a job, $\frac{1}{\theta}$, [the rate of occurrence] in each age group is taken to be constant i.e. the distribution of waiting time to have a job is exponential with the p.d.f. given in equation (2.7).
4. Length of non-working life in a certain age group is measured by the gross years of non-working life, where:

The gross years of non-working life, represents the difference between probable gross years and actual gross years of working life and actual gross years of working life is calculated to each age group as the product of the age-specific activity rate by the length of the age group interval.

3.2 Model Building

1. Length of non-working life is calculated as the difference between the length of the age interval and the actual gross years of working life.
2. Length of non-working life in a certain age-group is used as an estimate of the mean survival time in the unemployment state.
3. Hence, the moment estimator, $\hat{\theta}$, of the parameter θ can be obtained .
4. A random sample can be generated using the resulting estimated waiting time distribution. It must be censored at the end of the age interval i.e. we have a Type I censored sample.
5. The resulting sample can be used to make the required statistical inference.

3.1 Model Extensions.

1. The assumption of a constant hazard function is a very restrictive one, it is more realistic to assume that the instantaneous probability of having a job in each age group is increasing or decreasing and the model with increasing or decreasing hazard rate is more suitable.

2. The model may be extended by using a more refined measure to the length of non-working ages than the gross years of non-working life such as the net years of non-working life which represents the length of non-working years for a generation including persons whose working life is affected by death before they reach retirement age and hence takes into account both the level of economic activity and losses due to mortality. Also, measures obtained using working life tables can be used.

4. APPLICATION

The burden of unemployment in Egypt is increasing. The results of the 1996 census of Egypt revealed that the length of non-working life for males is higher among younger ages in all educational categories, but this measure is higher for females in middle ages because of increasing household work and caring for children in these ages. Also, length of non-working life is higher for females in all age groups for different educational categories .

Education is one of the most important factors affecting unemployment. The basic concern here is to investigate the effects of different educational categories on durations of unemployment through an application of the survival model suggested in Section 3 to the case of Egypt in 1996 .

In the present application, survival analysis for durations of unemployment is carried out for the following three educational categories:-

1.Less than intermediate [which will be denoted by the 1st educational category].

2.Intermediate [which will be denoted by the 2nd educational category].

3.University degree [which will be denoted by the 3rd educational category].

The present application concentrates on younger ages in which individuals begin to search for work after they complete their schooling . Hence, the age group 15-25 is studied for the first two educational categories and the age group 25-40 is studied for the university degree .

Separate analysis is necessary for males and females to represent sex differentials.

The application of the model is developed through the following steps :-

First :

using data of the most recent population census conducted in A.R.E. in 1996, the length of non-working life is calculated as follows :

(1) Age specific employment rates (A.S.E.R) are calculated, where the A.S.E.R in a certain age group [e.g. the jth age group] is defined as the number of employed individuals in the jth age group divided by the number of individuals at the mid point of the jth age group .

The data required for the calculation of the A.S.E.R are : -

- The distribution of individuals by age and educational status .
- The distribution of employed individuals by age and educational status .

(2) Actual gross years of working life is calculated to each age group as the product of the A.S.E.R by the length of the age group interval .

(3) Length of non-working life measured by the gross years of non-working life is then calculated as the difference between probable gross years and actual gross years of working life . This measure in the selected age groups is found to be 5.010, 6.150 and 1.740 years for males who received less than intermediate, intermediate and higher education respectively. The corresponding measure for females is found to be 9.680, 8.040 and 5.565 years.

Calculations are shown in tables (1) to (6) in the Appendix.

Second :

The calculated length of non-working life measured by the gross years of non-working life is used as an estimator of the mean survival time in the unemployment state, and we have the following moment estimators for θ :

$$\hat{\theta}_{1m} = 5.010 \ , \quad \hat{\theta}_{2m} = 6.150 \quad , \quad \hat{\theta}_{3m} = 1.740$$
$$\hat{\theta}_{1f} = 9.680 \ , \quad \hat{\theta}_{2f} = 8.040 \quad , \quad \hat{\theta}_{3f} = 5.565$$

where

θ_{km} is the mean survival time in the unemployment state for males in the k th educational category, and

θ_{kf} is the mean survival time in the unemployment state for females in the k th educational category

and k = 1, 2, 3 .

Having determined the mean survival time in the unemployment state, the corresponding waiting time distribution defined in equation (2.7) is determined, and we have the following estimated waiting time distributions

$$\hat{f}_{1m}(t) = 0.200\exp(-\frac{t}{5.010}) , \quad t \geq 0 \tag{4.1}$$

$$\hat{f}_{2m}(t) = 0.163\exp(-\frac{t}{6.150}) , \quad t \geq 0 \tag{4.2}$$

$$\hat{f}_{3m}(t) = 0.575\exp(-\frac{t}{1.740}) , \quad t \geq 0 \tag{4.3}$$

$$\hat{f}_{1f}(t) = 0.103\exp(-\frac{t}{9.680}) , \quad t \geq 0 \tag{4.4}$$

$$f_{2f}(t) = 0.124\exp(-\frac{t}{8.040}) , \quad t \geq 0 \tag{4.5}$$

$$\hat{f}_{3f}(t) = 0.180\exp(-\frac{t}{5.565}) , \quad t \geq 0 \tag{4.6}$$

where

$f_{km}(t)$ is the waiting time distribution in the unemployment state for males in the k th educational category , and

$f_{kf}(t)$ is the waiting time distribution in the unemployment state for females in the k th educational category,

and k = 1,2,3.

Third:

IMSl computer program is used to generate type I censored samples of size n =150 for the resulting estimated waiting time distributions defined in equations (4.1) to (4.6).

The resulting samples are used to study the waiting time distribution of unemployment in the pre-determined three educational categories and to compare the effects of these educational categories upon survival times in the unemployment state. Also, quantities such as the proportion of unemployed individuals that will have a job within a specified time can be estimated.

Survival probabilities in the unemployment state for the three educational categories in ages under study are illustrated in table (4.1). From table (4.1), we found that, the survival function in the unemployment state is a decreasing one. The probabilities of survival in the unemployment state are higher in younger ages for all educational categories and these probabilities are higher for females than for males in all age groups. The very highest probabilities of survival in the unemployment state are found among males with intermediate education and among females with less than intermediate education.

Point estimation and confidence intervals for the mean time in the unemployment state, $\hat{\theta}$, and the hazard function , $\frac{1}{\theta}$, for the three educational categories are displayed in Table (4.2) .

Table (4.1)

Survival Probabilities $\left[S(t_0) = \exp(-\frac{t_0}{\theta}) \right]$ **in the Unemployment State for the Three Educational Categories in Ages Under Study.**

Educational Category / Age	1st		2nd		3rd	
	M	F	M	F	M	F
15	1	1	1	1		
16	0.819	0.902	0.850	0.883		
17	0.671	0.813	0.722	0.780		
18	0.550	0.734	0.614	0.689		
19	0.450	0.662	0.522	0.608		
20	0.369	0.597	0.444	0.537		
21	0.302	0.538	0.377	0.474		
22	0.247	0.485	0.320	0.419		
23	0.203	0.438	0.272	0.370		
24	0.166	0.395	0.232	0.327		
25					1	1
26					0.563	0.836
27					0.317	0.698
28					0.178	0.583
29					0.100	0.487
30					0.057	0.407
31					0.032	0.340
32					0.018	0.284
33					0.010	0.238
34					0.006	0.198
35					0.003	0.166
36					0.002	0.139
37					0.001	0.116
38					0.001	0.097
39					0.000	0.080
40					0.000	0.000

Table (4.2)

Point Estimation and Confidence Intervals for the Mean Time in the Unemployment State, $\hat{\theta}$, and the Hazard Function, $\frac{1}{\theta}$, Calculated Using the Generated Type I Censored Samples

Estimated Value / Educational category		$\hat{\theta}$	95% confidence interval for θ	$1/\theta$	95% confidence interval for $1/\theta$
1st	M	5.077	$P[4.205<\theta<5.965]=0.95$	0.197	$P[0.168<1/\theta<0.238]=0.95$
	F	9.545	$P[7.626<\theta<11.464]=0.95$	0.105	$P[0.087<1/\theta<0.131]=0.95$
2nd	M	6.237	$P[5.112<\theta<7.362]=0.95$	0.160	$P[0.136<1/\theta<0.196]=0.95$
	F	8.154	$P[6.580<\theta<9.728]=0.95$	0.123	$P[0.103<1/\theta<0.152]=0.95$
3rd	M	1.720	$P[1.445<\theta<1.995]=0.95$	0.581	$P[0.501<1/\theta<0.692]=0.95$
	F	5.643	$P[4.705<\theta<6.581]=0.95$	0.177	$P[0.152<1/\theta<0.213]=0.95$

REFERENCES

Addison, J. T. and Portugal, P. (1992): Advance notice and unemployment: New evidence from the 1988 displaced worker survey. *Industrial and Labor Relations Review, Vol.45, No.4, pp. 645-664.*

Andrew, D. and Alan, D. (1992): Lone mothers, family credit and paid work .*Fiscal Studies, Vol. 13, No.1, pp.1-21.*

Bain, L. J. (1978): Statistical analysis of reliability and life-testing models. Theory and Methods. *Marcel Dekker, INC. New York and Basel.*

Balch, B. W. and Bichaka, F. (1997): Declining government involvement in job search and the rise of self-placement: A macroeconomic explanation. *American Business Review, Vol. 15,No.1,PP. 43-48.*

Datta, G.S. , Lahiri, P. Maiti, T. and Lu, k. L.(1999) : Hierarchical Bayes estimation of unemployment rates for the states of the U.S. . *J.A.S.A. , Vol. 94, No. 448, Applications and case studies .*

El-Biblawi, H. (1990): Unemployment and the loss of economically active years of life in Egypt; *Cairo Demographic center (CDC). Research Monograph Series, No. 19.*

El-Waffa, I, Z. A. (1993): Unemployment of the labour force and loss of active years of life in Egypt ,1988. *Cairo Demographic Center (CDC) DOC. CDC / S.93/oT.2-4*

Firth, D. and Payne, C. (1999): Efficiacy of programmes for the unemployed: discrete time modelling of duration data from a matched comparison study. *J. R. statist. Soc. A, 162, part 1, pp.111-120.*

Fusun, G. (1992): New evidence on whether unemployment and out of the labor force are distinct states. *Journal of Human Resources, Vol. 27, No. 2, pp. 329-361.*

Gourieroux, C. and Scaillet, O. (1997): Unemployment insurance and mortgages. *Insurance: Mathematics and Economics, Vol. 20, No.3, PP. 173-195.*

James, J. A. (1995): Reconstructing the pattern of American unemployment before the First World War. *Economica, Vol. 62, No. 247, PP.291-311.*

Lancaster, T. (1979): Econometric methods for the duration of unemployment. *Econometrica, Vol. 47, No. 4, pp. 939-956.*

Miles, C. (1993): Is unemployment insurance addictive ? Evidence from the benefit durations of repeat users. *Industrial and Labor Relations Review, Vol. 47,No.1, pp. 62-72.*

Poterba, J. M. and Summers, L, H. (1995): Unemployment benefits and labor market transitions. A multinomial logit model with errors in classification . *Review of Economics and Statistics, Vol. 77; No. 2,pp. 207-216.*

APPENDIX

Table (1)

Gross Years of Non-Working Life for Males with Less Than Intermediate Education**

Age Groups (1)	Population Total* (2)	Employed individuals* (3)	A.S.E.R.	Actual Gross Years of Working Life	Gross Years of non-Working Life
15-	3916122	1954325	0.499	4.990	5.010
25-	3239062	3112265	0.961	14.415	0.585
40-	2845339	2719157	0.956	14.340	0.660
55-60	625040	569299	0.911	4.555	0.445

* Source: CAPMAS " Census of population in A.R.E. in 1996 ", Cairo, 1998, no. 1102/1998.

** Less than Intermediate [include: illiterate, read and write, less than intermediate].

Table (2)

Gross Years of Non-Working Life for Males with Intermediate Education

Age Groups (1)	Population Total* (2)	Employed individuals* (3)	A.S.E.R.	Actual Gross Years of Working Life	Gross Years of non-Working Life
15-	204760	787634	0.385	3.850	6.150
25-	1603805	1281717	0.799	11.985	3.015
40-	553513	528079	0.954	14.310	0.690
55-60	67627	59315	0.877	4.385	0.615

* Source: The same as in Table (1)

Table (3)

Gross Years of Non-Working Life for Males with a University Degree

Age Groups (1)	Population Total* (2)	Employed individuals* (3)	A.S.E.R.	Actual Gross Years of Working Life	Gross Years of non-Working Life
15-	155064	100049	0.645	6.450	3.550
25-	824223	728857	0.884	13.260	1.740
40-	499670	483381	0.967	14.505	0.495
55-60	68557	61165	0.892	4.460	0.540

* Source: The same as in Table (1)

Table (4)

Gross Years of Non-Working Life for Females with Less Than Intermediate Education**

Age Groups (1)	Population Total* (2)	Employed individuals* (3)	A.S.E.R.	Actual Gross Years of Working Life	Gross Years of non-Working Life
15-	3848759	124442	0.032	0.320	9.680
25-	4366953	151311	0.035	0.525	14.475
40-	3374770	132247	0.040	0.600	14.400
55-60	655568	20480	0.031	0.155	4.845

• Source: The same as in Table (1).

** The same as in Table (1).

Table (5)

Gross Years of Non-Working Life for Females with Intermediate Education

Age Groups (1)	Population Total* (2)	Employed individuals* (3)	A.S.E.R.	Actual Gross Years of Working Life	Gross Years of non-Working Life
15-	1629851	319414	0.196	1.960	8.040
25-	1102457	397400	0.361	5.415	9.585
40-	252815	151603	0.600	9.000	6.000
55-60	24620	10703	0.435	2.175	2.825

* Source: The same as in Table (1).

Table (6)

Gross Years of Non-Working Life for Females with a University Degree

Age Groups (1)	Population Total* (2)	Employed individuals* (3)	A.S.E.R.	Actual Gross Years of Working Life	Gross Years of non-Working Life
15-	132239	70932	0.536	5.360	4.640
25-	477322	300079	0.629	9.435	5.565
40-	186590	134522	0.721	10.815	4.185
55-60	14145	8188	0.579	2.895	2.105

* Source: The same as in Table (1)

&&&&&&&&&&&&&&&&