age distributions of the emigrants from Greece

by
George N. Tziafetas

Lecturer
National Technical University of Athens

1. introductory remarks

In many social phenomena the duration of some particular activities is a useful measure of behavior. Thus, for example, the length of human life or specially the age at migration as a measure of staying by an individual in a region is of fundamental interest to the researchers. Of course, such phenomena are subjected to a high degree of variability and furthermore to an appropriate stochastic analysis. From this point of view, we will try to study the migration process by comparing frequency distributions predicted by model-building with those observed in practice.

The length of life staying in a region or more generally the duration of staying in a social state[1] can be studied in many cases analogously as in other areas outside the social sciences. The best known and most developed is the reliability theory and the industrial life testing. In spite of having important differences between the social and industrial applications according to the types of the distributions and collections of the data,[2] we shall follow the terminology and notation given in the reliability theory.

The main variable T underlying to investigation denotes the age of the individuals migrated from a region. We assume that T gives the duration of staying in the concerned region, if an individual is born there. Apart from mentioned exceptions, T will be continuous. Under these conditions, we can draw up a table of useful quantities that are objects of our main interest.

(a) The distribution function F(t). This is the cumulative probability

$$F(t)=P(T\leq t)$$

an individual to stay in the concerned region until the age t.

(b) The survivor function G(t). This is the complement of the distribution function F(t).

(c) The migration-age density function f(t). This is an equivalent quantity to completed length of service density function (CLS-distribution) in life testing and it is defined as the density function associated with F(t), related to the survivor function by

$$f(t)=-\frac{dG(t)}{dt} \qquad (1.1)$$

(d) The number of emigrants in time interval (t, t+1) is defined by

$$n\left[G(t)-G(t+1)\right] \qquad (1.2)$$

where n denotes the total number of out-migrants.

1. As a well known example can be regarded the phenomenon of labour turnover.
2. Bartholomew (1973).

(e) The intensity function or failure rate $\lambda(t)^3$ can be defined as follows:

Prob {migration in (t, t+δt)/ survival to t}= λ(t)δt.

Furthermore we have

f(t)= Prob {migration in (t, t+δt)}
= Prob {survival to t} Prob {migration in (t, t+δt) /survival to t}.

Hence $$\lambda(t)=\frac{f(t)}{G(t)}=-\frac{d \log G(t)}{dt} \tag{1.3}$$

or $$G(t)=\exp\left\{-\int_0^t \lambda(x)dx\right\} \tag{1.4}$$

(f) The specific rate of migration r(t) at age t can be defined as follows:

$$r(t)=\frac{G(t)-G(t+1)}{\int_t^{t+1} G(x)dx} \tag{1.5}$$

2. age distribution of the temporary emigrants from Greece

Lane and Andrew in their notable contributions introduced the log-normal distribution to analyse the phenomenon of labour turnover. Afterwards their findings have been confirmed many times and have become a principal tool in manpower planning. Hence, we may also apply the same principle in the present context to deduce that the form of the age distribution tends to lognormality.

Let T_j denotes the length of life time spent at least by an individual in a subregion j. If $\{u_2, u_3,...,\}$ is a sequence of random variables with known joint distribution, then the migration-age T and the law states are

$$T_j=T_{j-1}\,u_j \quad (j=2,3,...,) \tag{2.1}$$

$$T=T_j\,u \tag{2.2}$$

where u is a random variable.

Equation (2.1) and (2.2) may be regarded as expressions of the way that previously influences the future migration. An individual who has been staying for a long time in a region will tend to stay a long time in future.

An immediate consequence of equation (2.1) is that

$$T=u\,T_1\prod_{i=2}^{j} u_i \qquad (j\geq 2)$$

or

$$\log T=\log T_1+\log u+\sum_{i=2}^{j} \log u_i \tag{2.3}$$

Provided that the joint distribution of the u's is such that the central limit theorem applies to the sum of their logarithms, it follows that logT will tend to normality as j increases. Thus, the model predicts that a new emigrant who has been in several subregions in the past will have a migration-age density function which is approximately log-normal.

We assume further that u_i (i=2, 3, ...,) has the same joint distribution for all individuals of the population from which the migrants come. Under this condition, we can draw an important conclusion. Suppose that we classify all migrants according to the number of regions where they stayed in the past. Then, within these groups the model predicts that the standard deviation σ of logT should be constant. On the other hand, the mean m of logT depends on the characteristics of the individuals through $\log T_1$. Even if the groups are not homogeneous with respect to the number n of previous subregions the conclusion may still apply, because if the factors u_i have mean near to one then $\log u_i$ will have mean near to zero. So that σ will depend only slightly on n. Provided that n does not vary too much, the shape of the overall distribution is not likely to be seriously distorted by the effect which variation in n will have on σ.

The theory can go further, however, by supposing that these initial characteristics reflected in the quantities T_1 can be modified by subsequent stay in another subregion. Even, if this explanation is only partly true, it has an important practical consequence. Equation (2.3) shows an important role played by length of life time spent in the first subregion on subsequent lengths of stay. This suggests that if particular care is taken to ensure satisfaction in a person's first stay then their subsequent rate of migration will be reduced. We note here that the existing data do not provide enough information to enable a satisfactory test about such a hypothesis to be made.

In order to apply these considerations to the temporary emigrants of Greece, first we need an estimation of the parameters. Thus, from the knowledge that T has a two parameter log-normal distribution, one can obtain estimates of m and σ with optimum properties by considering the distribution W= log T. The maximum-likelihood estimator $\hat{m}$ of m is the sample mean given by

$$\frac{\sum_{k=1}^{v} f_k\, w_k}{v}=\frac{1}{v}\sum_{k=1}^{v} f_k \log t_k \tag{2.4}$$

where t_k is the kth ordered observable sample with frequency f_k from a log-normal distribution and v is

3. The intensity rate is known in the literature of reliability theory as hazard rate. In actuarial statistics the intensity rate goes under the name of mortality.

the sample size. The maximum-likelihood estimator $\hat{\sigma}^2$ of σ^2 is

$$\frac{1}{v}\sum_{k=1}^{v} f_k (\log t_k)^2 - \frac{1}{v}\Big[\sum_{k=1}^{v} f_k \log t_k\Big]^2 \qquad (2.5)$$

On the basis of observed data from tables (2.1), (2.2), (2.3) and (2.4) we calculated the parameters for the pertinent groups of the emigrants according to the marital status and for the total distribution. An estimation of the standard deviation is about the same equal to $\hat{\sigma}= 0.3$. On the other hand we have estimated different values for the mean of the three groups. The fitted distributions and the whole results are illustrated in the mentioned tables with the observed distributions.

Having the form of the empirical distribution we can construct a second model on the basis of gamma distributions. Thus, let us assume now that the kth individual migrates from a region at time $T^{(k)}$ and the time T_i between ith and (i—1)th migration has an exponential distribution with a constant parameter λ. This is a similar consideration to the constant failure rate. The probability density function of the random variable $T^{(k)}$ is defined by the k-fold convolution of the T_is as [4]

$$f_{T(k)}(t) = [f_{T_i}(t)]^k = \int_0^t [f_{T_i}(t-s)]^{k-1} f_{T_i}(s)\,ds \qquad (2.6)$$

where

$$f_{T_i}(t) = \lambda e^{-\lambda t} \qquad (\lambda > 0,\ t \geq 0) \qquad 2.7$$

Accordingly

$$[f_{T_i}(t)]^2 = \int_0^t f_{T_i}(t-s)\, f_{T_i}(s)\,ds =$$

$$= \int_0^t \lambda^2 e^{-(t-s)} e^{-ts}\,ds = \frac{\lambda^2 t e^{-t\lambda}}{\Gamma(2)}$$

In general, it can be deduced that

$$f_{T(k)}(t) = [f_{T_i}(t)]^k = \frac{\lambda^k}{\Gamma(k)} t^{k-1} e^{-\lambda t} \qquad (2.8)$$

which is the gamma probability density junction for $k \geqslant 1$.

In order to apply the gamma distribution given by (2.8) to the temporary emigrants age-distribution in Greece, first we need an estimation of the two parameters. An efficient method of parameter estimation is that of maximum-likelihood. Thus, the maximum likelihood equations reduce to

4. Feller, 1966, p. 45.

$$\left(\frac{\hat{1}}{\lambda}\right) = \frac{\sum_{k=1}^{v} f_k t_k}{v\hat{k}} \qquad \text{and}$$

$$2\hat{k} \simeq \log\Big(\sum_{k=1}^{v} \frac{f_k t_k}{v}\Big) - \frac{1}{v}\sum_{k=1}^{v} f_k \log t_k\Big]^{-1} + \frac{1}{3} \qquad (2.9)$$

where t_k is the kth ordered observable sample with frequency f_k and v is the sample size.

Alternatively, the method of moments can be employed for estimating λ and k. Moment estimators, given by

$$\left(\frac{\tilde{1}}{\lambda}\right) = \frac{\sum_{k=1}^{v} f_k t_k}{v\tilde{k}} \qquad \tilde{k} = \frac{\sum_{k-1}^{v} f_k t_k}{v\sum_{k=1}^{v} f_k (t_k - \bar{t})^2} \qquad (2.10)$$

are obtained by equating $k(1/\lambda)$ and $k(1/\lambda)^2$ to the sample mean and the variance respectively. These estimators have a lower asymptotic efficiency than the corresponding maximum-likelihood estimators.

On the basis of the observed data from the tables (2.1), (2.2), (2.3) and (2.4) we calculated for the concerned three groups of the emigrants according to the marital status and for the total distribution of λ and k by the method of moments.

3. age distribution for the permanent emigrants from Greece

Silcock (1954) published two empirical CLS-distribution obtained in his studies for different firms. His first model was based on a constant loss intensity that leads to an exponential CLS-distribution. It is doubtful whether Silcock's interpretation is justified, since the loss intensity could centainly depend on factors internal to the firm, including the individuals length of service. The second model proposed by Silcock is a generalization of the first. It retains the simple assumption of constant loss rate for individuals but this rate is now supposed to vary in the population from which employees are drawn. This is a very plausible hypothesis. Individuals differ almost from all other aspects of their behavior. In order to apply these considerations Silcock introduced an exponential CLS-distribution supposing that λ is a random variable with a gamma distribution and calculated the compound distribution.

Now, for our application to the permanently emigrants from Greece, it is simpler to suppose that a constant loss intensity can be regarded at least for a group of the emigrants, who have a commom be-

TABLE 2.1. *Observed and Fitted Age-distribution of the Temporary Emigrants from Greece: 1975 (males)**

Age (years)	Actual number of out-migrants	Log-normal fit	Gamma fit
0-4	—	0	6
5-9	12	4	46
10-14	57	360	794
15-19	9070	3298	3770
20-24	11259	9507	8800
25-29	16846	14438	13145
30-34	10822	15086	14485
35-39	9826	12442	12814
40-44	7784	8847	9606
45-49	5814	5696	6327
50-54	3094	3435	3756
55-59	1761	1982	2048
60-64	831	1109	1041
65-69	280	609	498
70-74	55	330	226
75-79	12	178	98
80-84	1	95	41
>85	—	111	26
Total	77527	77527	77527
		$\hat{m}=1,90629$	$\hat{k}=10,286$
		$\hat{\sigma}=0,31093$	$\hat{\lambda}=0,687$

* Source: *Statistical Yearbook of Greece* (1975).

TABLE 2.2. *Observed and Fitted Age-distribution of the Temporary Emigrants from Greece: 1975 (males-single)**

Age (years)	Actual number of out-migrants	Log-normal fit	Gamma fit
0-4	—	0	0
5-9	12	0	21
10-14	57	245	692
15-19	8842	3404	4013
20-24	10277	9706	8806
25-29	11585	11402	10380
30-34	4489	7950	7977
35-39	2092	4111	4494
40-44	1007	1693	2002
45-49	597	627	742
50-54	254	209	238
55-59	149	75	67
60-64	63	20	17
65-69	23	8	4
70-74	5	3	1
75-79	1	1	0
80-84	1	0	0
>85	0	0	0
Total	39454	39454	39454
		$\hat{m}=1,71163$	$\hat{k}=14,089$
		$\hat{\sigma}=0,28000$	$\hat{\lambda}=0,406$

* Source: *Statistical Yearbook of Greece* (1975).

TABLE 2.3. *Observed and Fitted Age-distribution of the Temporary Emigrants from Greece: 19[illegible] (males-married)**

Age (years)	Actual number of out-migrants	Log-normal fit	Gamma fit
0-4	—	—	—
5-9	—	—	—
10-14	—	0	0
15-19	224	7	28
20-24	962	427	548
25-29	5197	2813	2617
30-34	6236	6365	5771
35-39	7644	7970	7810
40-44	6709	7146	7576
45-49	5111	5197	5774
50-54	2794	3335	3669
55-59	1585	1961	2023
60-64	751	1095	996
65-69	248	592	446
70-74	48	314	185
75-79	11	163	72
>80	0	175	45
Total	37560	37560	37560
		$\hat{m}=1,81702$	$\hat{k}=11,168$
		$\hat{\sigma}=0,31611$	$\hat{\lambda}=0,576$

* Source: *Statistical Yearbook of Greece* (1975).

TABLE 2.4. *Observed and Fitted Age-distribution of the Temporary Emigrants from Greece: 1975 (males -divorced)**

Age (years)	Actual number of out-migrants	Log-normal fit	Gamma fit
0-4	—	—	—
5-9	—	—	—
10-14	—	0	0
15-19	1	0	0
20-24	19	7	6
25-29	57	39	40
30-34	86	76	77
35-39	80	88	94
40-44	55	74	81
45-49	53	51	54
50-54	35	32	30
55-59	12	18	14
60-64	5	11	6
65-69	1	6	2
70-74	1	3	1
>75	0	0	0
Total	405	405	405
		$\hat{m}=1,7787$	$\hat{k}=11,7147$
		$\hat{\sigma}=0,3242$	$\hat{\lambda}=0,5313$

* Source: *Statistical Yearbook of Greece* (1975).

havior depended on their educational obligations. On the other hand we assume that the permanently emigrant in higher ages can be distributed by ages according to a log-normal or gamma law in the same way as the temporary emigrants. If the associated probabilities are a_1 and $1-a_1$ respectively the distribution will be according to the first model

$$f(t)=a_1\lambda_1 e^{-\lambda_1 t}+(1-a_1)\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left\{-\frac{(\log t-m)^2}{2\sigma^2}\right\}$$

and by the second model

$$f(t)=a_1\lambda_1 e^{-\lambda_1 t}+(1-a_1)\frac{\lambda_2^k}{\Gamma(k)}x^{k-1}e^{-\lambda_2 x}$$

These models are two-term mixed distribution and are fitted to the data given by the National Statistical Service of Greece. For estimating the parameters of such mixed models, several graphical and theoretical methods have been proposed. For our models we assume that a_1= 0,1 or 10% of the total emigrants will be distributed by age according to a constant intensity failute rate. This group must be the young individuals of the migrated population. From this reason we can take an estimation of the parameter λ_1 having only the first age group, that is distributed according to the exponential law. Thus, we have $\hat{\lambda}_1$= 0.7199.

In table (3.1), we illustrated the whole results about the fitted log-normal and gamma distributions having estimators for the parameters given by the relationships (2.4), (2.5) and (2.9).

According to the marital status of the emigrants we calculated again estimators for the parameters of the distributions considering the same hypothesis about the distribution of young individuals. Thus, for unmarried permanent emigrants, we assume that 10% or 1171 individuals are distributed according to the exponential law and the rest according to the log-normal or gamma law. For the two other groups we have not any problem because we take emigrants only in higher ages. The whole results are illustrated in table (3.2), (3.3) and (3.4).

Although the models described above are in good agreement with the data, it cannot be concluded from the foregoing analysis that mixed models provide a true explanation of the migration process. On the other hand the models suggest that we can find ways for discriminating between emigrants according to their propensity to leave. But, this is another problem more concerned to a further sociological analysis.

TABLE 3.1. *Observed and Fitted Age-distribution of the Permanent Emigrants from Greece: 1975 (males)**

Age (years)	Actual number of out-migrants	Exponential fit	Log-normal fit	A mixed model	Gamma fit	B mixed model
0-4	601	601	0	601	0	601
5-9	532	293	10	303	36	329
10-14	353	142	211	353	274	416
15-19	1326	69	848	917	795	864
20-24	1590	34	1535	1569	1371	1405
25-29	1954	16	1817	1833	1706	1722
30-34	1364	8	1689	1697	1710	1718
35-39	1337	4	1357	1361	1464	1468
40-44	996	2	995	997	1115	1117
45-49	726	1	689	690	775	776
50-54	339	1	459	460	500	501
55-59	183	0	298	298	393	393
60-64	150	0	191	191	176	176
65-69	79	0	121	121	97	97
70-74	48	0	76	76	52	52
75-79	14	0	48	48	27	27
80-84	5	0	30	30	14	14
>85	2	0	54	54	13	13
Total	11599	1171	10428	11599	10428	11599
			$\hat{m}$=1,86408		$\hat{k}$=7,5840	
		$\hat{\lambda}_1$=0,7199	$\hat{\sigma}$=0,37691		$\hat{\lambda}_2$=0,9097	

* Source: *Statistical Yearbook of Greece* (1975).

TABLE 3.2. *Observed and Fitted Age-distribution of the Permanent Emigrants from Greece: 1975 (males-single)**

Age (years)	Actual number of out-migrants	Exponential fit	Log-normal fit	A mixed model	Gamma fit	B mixed model
0-4	601	601	0	601	0	601
5-9	532	293	18	311	38	331
10-14	353	142	325	467	332	474
15-19	1301	69	937	1006	851	920
20-24	1339	34	1144	1178	1115	1149
25-29	917	16	903	919	958	974
30-34	273	8	559	567	620	628
35-39	144	4	304	308	327	331
40-44	92	2	153	155	148	150
45-49	48	1	74	75	59	60
50-54	19	1	35	36	22	23
55-59	11	0	16	16	7	7
60-64	13	0	7	7	2	2
65-69	5	0	3	3	1	1
70-74	1	0	1	1	0	0
>75	2	0	1	1	0	0
Total	5151	1171	4480	5651	4480	5651
			$\hat{m}=1,57484$		$\hat{k}=9,5317$	
		$\hat{\lambda}_1=0,7199$	$\hat{\sigma}=0,33355$		$\hat{\lambda}_2=0,5345$	

* Source: *Statistical Yearbook of Greece* (1975).

TABLE. 3.3. *Observed and Fitted Age-distribution of the Permanent Emigrants from Greece: 1975 (males-married)**

Age (years)	Actual number of out-migrants	Log-normal fit	Gamma fit
0-4	—	—	—
5-9	—	—	—
10-14	—	0	0
15-19	24	3	16
20-24	250	115	158
25-29	1023	581	554
30-34	1071	1101	997
35-39	1172	1238	1188
40-44	895	1034	1072
45-49	662	722	792
50-54	318	452	502
55-59	167	264	283
60-64	129	148	145
65-69	64	80	69
70-74	40	43	31
75-79	9	23	13
80-84	4	12	5
>85	0	12	3
Total	5828	5828	5828
		$\hat{m}=1,77419$	$\hat{k}=9,4379$
		$\hat{\sigma}=0,33010$	$\hat{\lambda}=0,6592$

* Source: *Statistical Yearbook of Greece* (1975).

TABLE 3.4. *Observed and Fitted Age-distribution of the Permanent Emigrants from Greece: 1975 (males-divorced)**

Age (years)	Actual number of out-migrants	Log-normal fit	Gamma fit
0-4	—	—	—
5-9	—	—	—
10-14	—	—	—
15-19	1	0	0
20-24	1	1	0
25-29	9	6	2
30-34	18	28	12
35-39	17	13	22
40-44	7	9	19
45-49	10	6	10
50-54	1	3	3
55-59	1	2	1
60-64	2	1	0
65-69	1	0	0
70-74	1	0	0
>75	0	0	0
Total	69	69	69
		$\hat{m}=1,77536$	$\hat{k}=23,731$
		$\hat{\sigma}=0,31435$	$\hat{\lambda}=0,254$

* Source: *Statistical Yearbook of Greece* (1975).

REFERENCES

1. Bartholomew, D.J., *Stochastic Models for Social Processes*, John Wiley (1973).
2. Feller, W., *An Introduction to Probability Theory and Its Applications*, vol. II, John Wiley (1966).
3. Lane, K. F. and J.E. Andrew, «A Method of Labour Turnover Analysis», *J.R. Stat. Soc.*, A 118, 296-323 (1955).
4. Silcock H., «The Phenomenon of Labour Turnover», *J. R. Stat. Soc.*, A 117, 429-440 (1954).
5. Tziafetas, G. «Zur Anwendung der Theorie der stochastischen Prozessen in der Sozialwissenschaft und insbesondere in der Demometrie», *Greek Rev. of Social Research*, no. 26-27 (1976).