The use of Coxian Phase-type Distributions to Model the University Student Progression and Drop out Rates: the case of an italian university.

Adele H., Marshall
Centre for Statistical Science and Operational Research (CenSSOR), Queen's University Belfast
Belfast, Northern Ireland. Uk
E-mail: a.h.marshall@qub.ac.uk

Mariangela, Zenga
Department of Quantitative Methods for Economics and Business, University of Milano-Bicocca
Via Bicocca degli Arcimboldi, 8
20126, Milano, Italy
E-mail: mariangela.zenga@unimib.it

Sabrina, Giordano; Filippo, Domma
Department of Economics and Statistics, University of Calabria,
Arcavacata di Rende, Cosenza, Italy.
E-mail: sabrina.giordano@unical.it; f.domma@unical.it

1.Introduction

The dynamic analysis of university education has been developing systematically for a long time. There is thriving literature about the educational system helping to drive towards the economic politics of a nation. Even if it means the university, the notion of *University System* can be different in a European view. In fact, in the European countries there are different models of university organisation that have been changing in the last decade. The reason for the changes is often represented by improving national social and economic factors. For the Italian situation, there was a radical university reform in 1999 (L.4/1999) that organized higher educational studies in compliance with the principles of the European agreements of Bologna (Declaration of 18-19 June 1999). Before this reform, the higher education in Italy was given by the degree with a variable legal duration (from four or six years according to the faculty). Differently from the other European countries, there was not an intermediate level of education. In 1998, more than 1,600,000 students attended university (against 450,000 in 1967 and 1,000,000 in 1978), but only 15% of graduated students took the degree in the legal time and the other had a *fuori corso*[1] position. The very problem of the Italian university system was the students leaving the university in the first two years of the university career. The 1999 reform had the following aims:

- to shorten the period taken to graduate and to eliminate the Italian phenomena of *fuori corso*;
- to reduce the number of students not completing courses;
- to reach a greater homogeneity within the university systems;
- to favour a greater mobility, both nationally and internationally, of students through the introduction of the credit system, and,
- to improve links between methodology and culture and an effective professional training.

According to the European model, university studies are now organized in 3 levels.
As shown in the paper given by the OECD organization, *Education at Glance 2004*, in Italy, the university enrollment increased by 8% between 1995 and 2002. However, despite this increase, educational tertiary completion is very low, with only 23% of an age cohort completing a first university degree, compared with an OECD average of 32% and 45% in Australia and Finland respectively. At 50% (44% for men and 57% for women), the university entry rate is now at the OECD average level. But almost 60% of those who enroll in university never

[1] A *fuori corso* student takes more than the degree legal time. They will have paid the university taxes and remained officially registered as a student in university for several years. They can spend more then 10 years in the university.

complete with a qualification. This data represents the highest drop-out rate among OECD countries (on average, the drop-out is 30%). This situation can be seen in Figure 1. This suggest some level of mismatch between the needs and aspiration of students and what current university programmes offer. It is noteworthy that the drop-out rate is much lower in programmes with a duration of 3- 4 years than in the old university system. That suggests that the move to a more flexible multi level qualification may lead to a gain in efficiency.

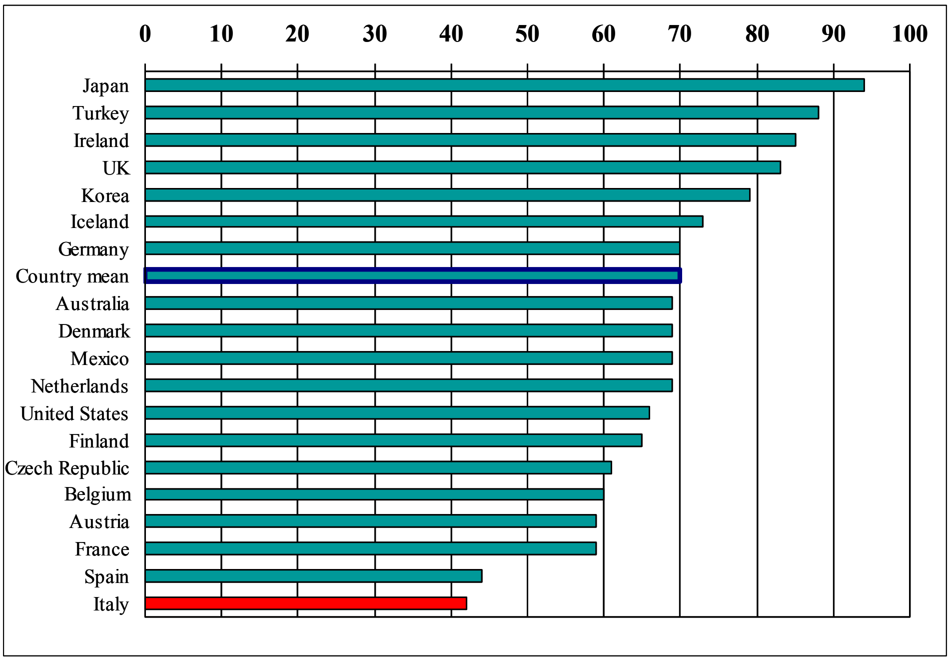

Fig 1: Number of graduates divided by the number of new entrants in the typical year of entrance by all programmes of the first level of university courses (2004).

In recent years, economists and sociologists are paying attention to the length of stay of students in university: it is analyzed particularly with respect to the drop out problem. In fact, the university education can be regarded as an investment [Schultz, 1963]. In particular, the economic performance of education has been studied, comparing the costs of education with the future benefits. When a student leaves the university before the degree, costs arise for the student's family, for the university as an institution and for the society. The university is affected by the withdrawal because the funds, given by the state, are determined also by performance indexes about the rate of not finishing the course of study [Smith, 2001]. The social costs regard the economic output loss: the gradated is more productive than the non-graduated and the society doesn't make a profit from the taxes of the missed graduate [Kiefer and Neumann, 1979].

This paper is organized as follows; first of all the Coxian phase type distribution will be introduced followed by analysis where student progression and drop-out rates are presented for data from the Milano-Bicocca university. Finally the Coxian phase type distribution is shown to provide a good representation of the length of stay of students at university.

2.Coxian phase-type distributions

A Coxian phase-type distribution $\{X(t);\ t \geq 0\}$ is a (latent) Markov chain in continuous time with states $\{1, 2, \ldots n, n + 1\}$, $X(0) = 1$, and for $i = 1, 2, \ldots n-1$

$$\text{prob}\{X(t+\delta t) = i+1 \mid X(t) = i\} = \lambda_i \delta t + o(\delta t) \tag{1}$$

and for i = 1, 2, ...n

$$\text{prob}\{X(t+\delta t) = n+1 \mid X(t) = i\} = \mu_i \delta t + o(\delta t). \tag{2}$$

Here states $\{1, 2, \ldots n\}$ are latent (transient) states of the process and state $(n+1)$ is the (absorbing) state. λ_i represents the transition from state i to state $(i+1)$ and μ_i the transition from state i to the absorbing state $(n+1)$ (as shown in Figure 2)

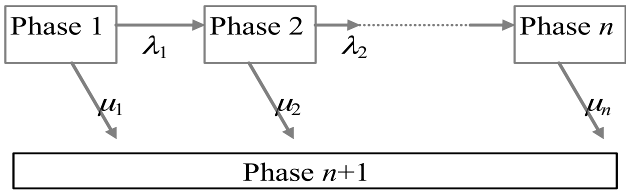

Fig. 2.:An illustration of the Coxian phase-type distribution.

The process begins in the first phase and may either progress through the phases sequentially or enter into the absorbing state (the terminating event - phase $n+1$). Such phases may then be used to describe stages of a process which terminates at some stage.

Cox and Miller (1965) develop the theory of Markov chains such as those defined by (1) and (2). The Coxian phase-type distribution is defined as having a transition matrix **Q** of the following form,

$$\mathbf{Q} = \begin{pmatrix} -(\lambda_1+\mu_1) & \lambda_1 & 0 & \ldots & 0 & 0 \\ 0 & -(\lambda_2+\mu_2) & \lambda_2 & \ldots & 0 & 0 \\ . & . & . & & . & . \\ . & . & . & & . & . \\ . & . & . & & . & . \\ 0 & 0 & 0 & \ldots & -(\lambda_{n-1}+\mu_{n-1}) & \lambda_{n-1} \\ 0 & 0 & 0 & \ldots & 0 & -\mu_n \end{pmatrix}, \tag{3}$$

where the λ_i's and μ_i's are from Cox and Miller's (1965) theory of Markov chains defined by (1) and (2). The probability density function (pdf) of the Coxian phase-type distribution for time t is given by

$$f(t) = \mathbf{p}\exp\{\mathbf{Q}t\}\mathbf{q} \tag{4}$$

where p is a $1 \times k$ vector of probabilities defining the initial transient phases and q is a $k \times 1$ vector of rates from the transient states to the absorbing state, as shown in (5) and (6) respectively.

$$\mathbf{q} = -\mathbf{Q1} = (\mu_1\ \mu_2 \ldots \mu_n)^{\mathrm{T}} \tag{5}$$

and

$$\mathrm{P} = (1\ 0\ 0\ \ldots 0\ 0\ ). \tag{6}$$

Previous research has successfully used Coxian phase-type distributions to represent survival times as the length of time until a certain event occurs, where the phases are considered to be stages in the survival and the absorbing, final stage, the event that occurs causing the individual or element to leave the system completely. For instance, this event could be a patient recovering from an illness, a patient having a relapse, an individual leaving a certain type of employment, a piece of equipment failing, or a patient dying. Faddy (1994) illustrates how useful the Coxian phase-type distributions are in representing survival times for various applications such as the length of treatment spell of control patients in a suicide study, the time prisoners spend on remand and the lifetime of rats used as controls in a study of ageing.

In particular, Faddy and McClean (1999) used the Coxian phase-type distribution to find a suitable distribution for the duration of stay of a group of male geriatric patients in hospital. They found that the phase-type distributions were ideal for measuring the lengths of stay of patients in hospital and showed how it was also possible to consider other variables that may influence duration. More recently, Marshall and McClean (2003) have demonstrated how the Coxian phase-type distribution can, unlike alternative approaches, adequately model the survival of various groups of elderly patients in hospital uniquely capturing the typical skewed nature of such survival data in the form of a Conditional phase-type model (C-Ph) which incorporates a Bayesian network of inter-related variables.

3. Student Progression and Drop out: the Milano-Bicocca example

The empirical analysis reported in this section is based upon a data set from administrative data on individual students from the Milano-Bicocca university. The data refer to all the students enrolled for the first time in the academic year 1997-1998, 1998-1999, 2000-2001 for the Economics, Law and Science faculties. The database consists of 7876 students. For every student the several variables are recorded: personal information collected at the time of application to university and information about the progress of each student. The personal student information includes for each student: age, sex, secondary school, note in secondary school, family income, place of living, cohort, date of enrolment, date of leaving.

For the students included in the data, student status and careers have also been observed for nine academic years. The students that are still enrolled after nine academic years are considered right censored in the analysis.

The intention of this paper is to analyze student progression in the reformed degree system and to estimate the influence of various factors on the probability that a student will progress successfully through their degree course. We propose to use the Coxian-Phase type distribution, to describe the length of stay (LoS, in days) of the students at the university, given the typical pattern of the student university life.

3.1. The data description

At the end of the nine observed academic years, 37.29% of students are still enrolled underlining that a very high amount of student find difficulty the complete their degree within the corect time. The dropout rate is 27.22% of students (n=2144, LoS Mean=832.2, LoS Median=593, LoS Variation coefficient=0.89), whereas 7.49% of the students changed the Faculty or the University in which they were studying (n=590, LoS Mean=1209.3, LoS Median=1080.5, LoS Variation coefficient=0.69). Only 28% of students takes the degree during the nine observed academic years (n=2205, LoS Mean=2473.24, LoS Median=2399, LoS Variation coefficient=0.19).

3.2. Fits of the Coxian phase-type distribution

Table 1 reports the estimation of the Coxian-Phase type distribution parameters using the EM-algorithm [Asmussen,et al. 1996]. From inspection of the results, it is apparent that a 15 phase Coxian distribution is the most suitable for the data. However, it is important to note that some of the parameters (μ_j) in the table are equal to zero therefore allowing these related phases to be aggregated together. In effect the phases with small parameter values are made redundant and it is only the most dominant phases with the large parameter values that are regarded as useful. This will also prevent an over-fitted model. By combining phases with small parameter estimates, a fitted distribution is achieved consisting of three phases. In general, this distribution can fit the real pattern of the students at the university. The results therefore indicate three substantial phases in which the absorption probability is higher then in the other phases. In fact, the student career goes through these three sequential steps. At the beginning, we imagine an explorative phase in which the students face up to the new way of studying, where some of them don't accept it and soon leave this challenge (the drop out students) Secondly the "survivors" go on staying in the system (that is studying at university), step by step, in different stages of maturity (sequential transition phases), and in these phases some understand that the efforts of so far aren't enough and they too leave (drop out students) before completing their course, other students instead continue their work, and reach the graduation. So, it is possible to consider these three phases as:

- an explorative phase: the first impact with the new environment resulting in some students leaving,
- an intermediate phase: a trial period in which we meet two kinds of students: the motivated who increase their interests by attending courses (and their human capital) and making exams and the hesitant second group of students who are choosing between leaving or continuing to pursue their studies,
- an outcome phase: in which the motivated students complete their degree.

We can imagine that among these big phases there are some natural transition stages of training for motivated students and of attempting to motivate less interested students.

Actually, in each phase it is possible to leave the process of study, but in some of these phases the probability is that small it is more likely that the students stay in the process than going out from it.

Phases	$\lambda_j/(\mu_j+\lambda_j)$	$\mu_j/(\mu_j+\lambda_j)$	Prob.
1	0.9929	0.0071	0.0071
2	1	0	0
3	0.8330	0.1670	0.1658
4	1	0	0
5	1	0	0
6	0.8860	0.1140	0.0943
7	1	0	0
8	1	0	0
9	1	0	0
10	1	0	0
11	1	0	0
12	1	0	0
13	1	0	0
14	0.4673	0.5327	0.3904
15	-	-	0.3424

Tab.1:The results of fitting the Coxian phase-type distribution.

Figure 3 reports the histogram of the length of stay alongside the estimated distribution. This provides a graphical evaluation of the fit of the estimated Coxian phase-type distribution. From inspection of Figure 3, it is apparent that the estimated Coxian phase-type distribution seems to cature the peaks in the real data set. This feature provides suficient evidence to wish to consider the Coxian phase-type distribution as a suitable candidate model to describe the survival times for the length of stay at university of the students.

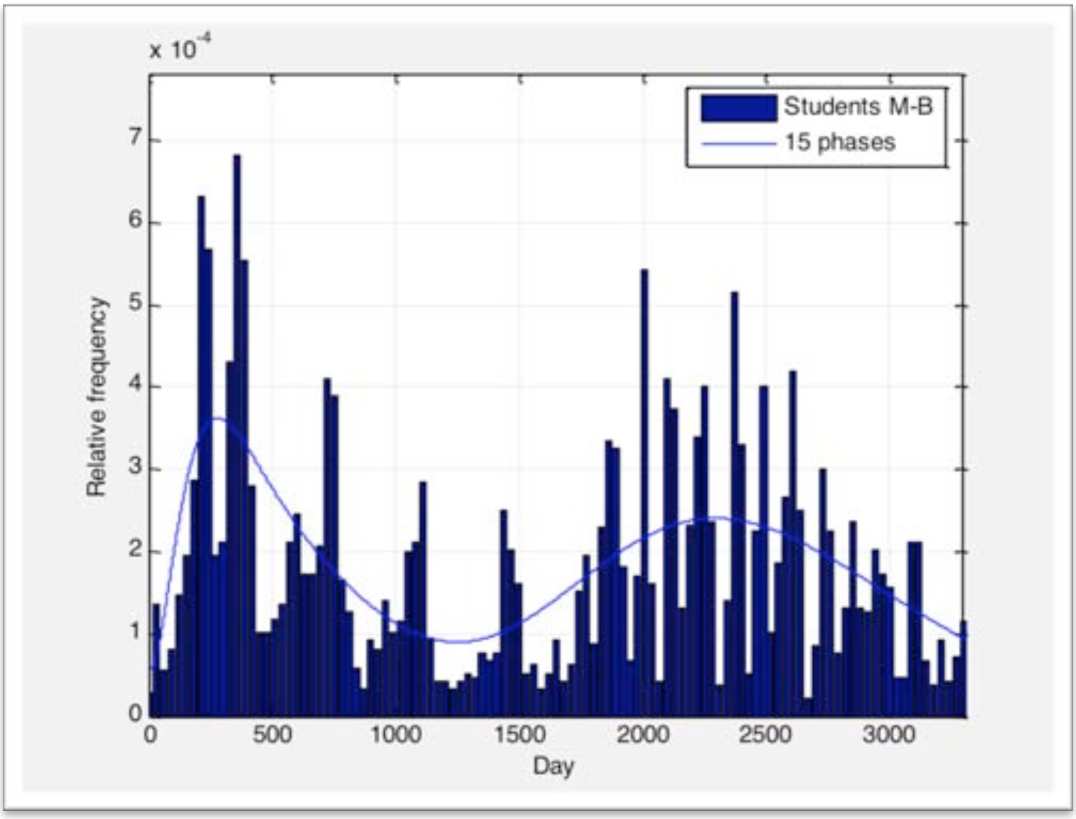

Fig.3:The empirical distribution and the estimated distribution.

4. Summary and Future Work

This work presents an innovative application of the Coxian phase-type distribution to the University student progression and drop out phenomena. In particular the model seems to identify three macro phases during the students length of stay at the university. Future work wishes to investgiate student covariates which could significantly affect the length of stay and incorporate these into the model possibly in the form of a Conditional phase-type distribution. In addition to this, it will be useful to determine two different probabilities of absorption for each event (degree and drop-out) from every phase, and consider how such probabilities are affected by the student covariates.

REFERENCES

1. S. Asmussen, O. Nerman, M. Olsson (1996.) Fitting phase-type distributions via the EM algorithm. *Scandinavian Journal of Statistics*, 23, 419–441.
2. D.R. Cox, H.D. Miller, The theory of stochastic processes. London: Chapman, 1965
3. M. Faddy, (1994) Examples of fitting structured phase-type distribution, *Applied Stochastic models and Data Analysis*, 10, pp 247-255.
4. M. Faddy, S.I. McClean, S.I. (1999) Analysing data on lengths of stay of hospital patients using phase-type distribution. *Applied Stochastic Models in Business and Industry* , pp 311–317
5. N. Kiefer, G. Neumann (1979), An empirical job-search model, with a test of the constant reservation wage hypothesis. *Journal of Political Economy* , 87, pp 89–107.
6. A.H. Marshall, S.I. McClean (2003) Conditional phase-type distributions for modelling patient length of stay in hospital. *International Transactions in Operational Research* 10, pp 565–576.
7. M. Neuts, Structured Stochastic Matrices of M/G/1 Type and Their Application (Marcel Dekker, NY, 1989).
8. T.W. Schultz, The economic value of education. New York: Columbia University Press, 1963.
9. J.P. Smith, A.N. Robin (2001) Dropping out of university: a statistical analysis of the probability of withdrawal for UK university students *Journal of the Royal Statistical Society. Series A,* Vol. 164, Part 2, pp 389-405.