Montserrat Hernández-LÓPEZ, PhD
José Juan Cáceres-HERNÁNDEZ, PhD
E-mail: jcaceres@ull.es
Departamento de Economía Aplicada y Métodos Cuantitativos
University of La Laguna

FORECASTING THE COMPOSITION OF DEMAND FOR HIGHER EDUCATION DEGREES BY GENETIC ALGORITHMS

Abstract. *A genetic algorithm is developed to forecast the relative presence of different university studies in the higher education demand in the field of economics and business/management as a whole. A selection operator is defined that assumes that the better the job opportunities associated with a specific university study, the higher the future demand for such a degree. A transition matrix takes other factors into account which may influence on the changes in demand. The proposed algorithm is applied to the original populations of students enrolled on 2005/2006 to 2007/2008 courses. Then, a new algorithm, whose elements are corrected to adjust the forecasts, is applied to obtain the forecast of the demand composition on the 2009/2010 course. This methodological proposal is shown to be able to provide the type of forecast which is very useful in policy making decisions in the recent process of building the European Higher Education Area.*

Key words: *higher education, forecasting, genetic algorithms.*

JEL Classification: C63, I23

1. Introduction

The forecasting of the whole demand for higher education is needed to make a decision about the amount of public economic resources to be assigned to the task to satisfy such a demand. This type of medium or long term forecast depends strongly on demography and income trends, and, also on access to information. However, in the short term, managers of universities must make decisions about the assignment of funds and, above all, human resources and physical space. Such decisions require an approximation to the demand for different university studies for the course following the one in which plans are presently being elaborated. To correct the imbalance between the supply of studies and the demands of students, changes in demand need to be predicted and the university planning policy should be designed taking present and future demands into account. From this point of view, the whole demand for university studies is not as relevant as its composition in terms of each of the studies which are supplied. The genetic algorithm proposed in this paper is a useful tool to this aim.

A student who has finished secondary education and is thinking about enrolling in a higher education center is able to brush aside those studies which are far from his preferences; but his vocation is not usually clear enough to determine the final choice. That choice depends on multiple factors (De Paola and Gioia 2012; Navarro and Casero 2012; Skatova and Ferguson 2014; Peró et al. 2015). Even so, in aggregated terms, vocations may be assumed to remain more or less the same in the short term and, therefore, changes in the number of students enrolled for different university studies may be assumed to be explained by the job opportunities associated to them, which also depend on student circumstances or characteristics.

Tribe (2003) remarks that universities are seeking to adapt their output of graduates to labor market demand. In fact, the European Higher Education Area is assigned the task of increasing this ability of university graduates and there is a general agreement between Spanish researchers about the presence of imbalances between higher education and social or economic demands (Salas 2004; García-Montalvo 2005; Fernández 2006; Mora 2008). Although workers with a higher education are less likely to be overeducated or undereducated compared to individuals with secondary education (Lassibille et al. 2001), a penalty in wage level may be a consequence of that education does not match the employment requirements (Dolado, Felgueroso and Jimeno 2000; Aguilar and García 2008). Technological change has brought about changes in relative demand for different university graduates, whose specific abilities are not equally valuable[1]. In a similar way, Grilli and Mealli (2008) find some evidence that there is a positive effect of some university studies on job opportunities.

However, students have usually got partial information about the labor market, and their motivations in choosing university studies need to be analyzed. As Hartog and Díaz-Serrano (2007) pointed out, individuals deciding on schooling are at best imperfectly aware of their abilities, the demands of the school curriculum, the probability to succeed, the nature of the job that may be obtained after completing their education and their position within the post-school earnings distribution that may be attained[2]. In spite of that, according to the human capital theory, higher education is an investment in the future. Businessmen select workers depending on their education and, therefore, investment in education improves the competitive advantage of workers in the labor market (Becker 1993; Barceinas et al. 2000;

[1] Using microdata from the Spanish sample of the 2007 Adult Education Survey, Nieto and Ramos (2013) conclude that non-formal education allows overeducated workers to acquire new abilities that improve their competence and allow them to earn higher wages. In accordance, Budría and Telhado-Pereira (2011) find that in most European countries the amount of wage dispersión within Education groups is substantially higher among college-educated workers than among less educated workers.

[2] In such a sense, Kucel and Vilalta-Bufí (2013) find that the effect of the level of education on job satisfaction becomes negative, probably due to the higher expectations of better educated workers.

Dolado et al. 2000; García-Montalvo 2005). If individuals are assumed to make decisions as investors, they will choose university studies associated with better job opportunities (Salas 2005). Following Latiesa (1989), the choice of a university program depends on the expectancies of academic success, the possibilities of getting a job, the personal preferences, the supply of the degrees, and the individual restrictions (Salas 1996; Salas 2005). Nonetheless, although some students indicate that vocation is the main factor in choosing a university degree, they highlight job opportunities as an important element in decision making. Furthermore, job opportunities may be behind the apparent choice by vocation, due to students associating degrees to specific jobs and life styles.

The aim of this paper is to forecast, by means of a genetic algorithm, the composition of the demand for a set of university studies as a function of the job opportunities associated to them. The selected studies are: a three year degree in Management, and two five year degrees, one in Economics, and one in Business Administration. All of these studies are supplied by the University of La Laguna, located in the Canary Islands (Spain). Thus, this paper analyses the choices of students between university studies in the field of economics and business/management, in such a way that these studies are close enough to assume that vocation is not the main reason of choice or, at least, that the changes in job opportunities are more relevant than the vocational changes in order to obtain the stated forecasts. From this point of view, the demand forecast should be based on measuring the effects of the chosen degree and other characteristics of the graduates on their job opportunities. Besides the degree, the graduates age and gender are two characteristics which could influence the probabilities of employment (Sáez and Rey 2000; Bellas 2001; Lassibille et al. 2001; Salas 2007).

The labor situation of graduates two years after finishing their degrees is also known. Three labor positions are distinguished: unemployed, with a temporary job or with a permanent job. The probabilities that a graduate is in one of these three positions two years after finishing the degree are evaluated by means of estimating a multinomial logit model. From the estimates of these probabilities a fitness function is defined that indicates the quality of access to the labor market and, in that sense, provides basic information for the forecasting procedure. Given that graduate characteristics are also available in the case of new students enrolled on the first course of a degree, a value of the fitness function can be assigned to each student by assuming that equal probabilities to get a specific type of job correspond to students with the same characteristics.

The fitness function is needed to define the selection operator, which is the main element of the performance of the genetic algorithm. The algorithm provides a forecast of the demand for different studies in terms of the percentage of new students enrolled for these studies for the course after the last one observed. In addition to the selection operator, transition matrices are also applied to correct errors in the forecasts obtained by the genetic algorithm with regards to the changes observed in the relative demand for the different degrees. In the following section the methodological elements of the genetic algorithm are developed. Next, the results obtained by applying the proposed algorithm are shown. Finally, concluding remarks are stated and some domains to which the investigation is intended to have implications are indicated.

2. Material and methods

The statistical tool proposed in this paper to forecast the composition of demand for higher education for a course is, as mentioned, a genetic algorithm (Holland 1975; Goldberg 1989). Some applications in economics are mentioned in Hernández-López and Cáceres-Hernández (2007). In some of them, genetic algorithms are applied as a statistical procedure that drives an optimization process. However, in this paper, the genetic algorithm, based on the natural selection principle that drives the transformations in different species populations, is applied as a mechanism that describes the dynamic process of transformation in the population of students enrolled on degrees in Management, Economics or Business Administration. Once the statistical procedure is developed, in this section it is also described both the population of graduates from which the probabilities to access the labor market are estimated, and the original populations to which the genetic algorithm is applied.

2.1. The design of the genetic algorithm

The algorithm modifies the original population of new students enrolled on these degrees on course t, and a final population of students enrolled on such degrees on the following course, $t+1$, is predicted. Individuals in these populations are identified by a structure or a bit chain that indicates their characteristics. The transformations that determine the relative frequency corresponding to each structure in the final population is guided by the selection operator in the first of two phases, and, once an intermediate population is obtained, a transition matrix operates in the second phase to transform the intermediate population into the final population.

2.1.1. Fitness function and selection operator

An increase is expected in the relative frequency of students enrolled on the degrees that provide better job opportunities. This is the basic assumption to define the selection operator. That is to say, the job opportunities associated with a structure determine its ability to adapt to labour market demands and, according to the natural selection principle, individuals better *adapted to the environment* are more likely to *survive* in such a way that their relative presence in future populations is also likely to be higher.

The so called fitness function, that measures the individual ability to adapt to labor circumstances, is defined from the estimates of probabilities that a student enrolled on a specific degree, besides other individual characteristics, gets a permanent job, gets a temporary job or remains unemployed in a time period after finishing the degree. To this aim, a multinomial logit model is estimated for the population of graduates on previous courses, identified by bit chains that indicate their characteristics. From these probabilities, the fitness function is defined as a function of individual characteristics. The ability to access to the labor market corresponding to individual i is assumed to be defined as

$$F_i{*} = 0P(Y_i = 1) + 3P(Y_i = 2) + 10P(Y_i = 3). \qquad (1)$$

where $Y_i = 1$, $Y_i = 2$ or $Y_i = 3$ if the individual i is unemployed, has got a temporary job or has got a permanent job, respectively. The weights assigned to probabilities $P(Y_i = j), j = 1,2,3$, are arbitrary values, but according to them the unemployed status is assumed to be the worst situation, whereas getting a temporary job reflects an improvement in job opportunities less than the one corresponding to the transformation of a temporary job into a permanent one.

As mentioned, the degree in Management is officially planned to take three years of study, whereas the degrees in Economics and Business Administration are planned to take five years. In this sense, a correction factor has been applied to the values of the fitness function corresponding to students enrolled on the degree in Management. According to the lengths of these studies, a 5/3 correction factor seems to be right. However, the actual length of the studies is at least a year longer than the officially planned lengths of the university studies analysed in this paper. In fact, data provided by the *Gabinete de Análisis y Planificación* of the University of La Laguna for the graduates on courses 2003/2004 to 2006/2007 suggest that the average lengths of the studies in Economics, Business Administration, and Management were, respectively, 6.4, 8.6 and 5.8 years. Furthermore, students have been assumed to believe that getting a specific job is more important than the time spent to get it. Then, a 4/3 correction factor has been applied so as not to produce an excessive distortion of the fitness function, defined as

$$F_i = \begin{cases} F_i{*} & , \quad \textit{Economics}, \; \textit{Business Administration} \\ \frac{4}{3}F_i{*} & , \quad \textit{Management} \end{cases} \qquad (2)$$

As the first step in the execution of the genetic algorithm, the corrected fitness function, F_i, is applied to the population of new students enrolled on course t. Once the structures observed in the original population are ordered according to individual characteristics, a selection operator is applied to obtain an intermediate population on the course $t+1$ in such a way that the probability that an individual is selected is proportional to the value of the fitness function for such an individual. If this selection operator were not able to capture the speed in the changes in the original population, another selection operator proportional to the square value of the fitness function could be applied.

In order to obtain an intermediate population of size n, the proportional selection operator applies the following procedure. If $\Omega_1 : \{I_{1,1},...,I_{1,r}\}$ denotes the set of individuals in the original population of size r and $W : \{1,...,n\}$ is defined as the set of n positions where the individuals copied from the original population are located, then the intermediate population $\Omega_2:\{I_{2,1},...,I_{2,n}\}$ is obtained through the selection operator $s(j)=I_{2,j}$, and is defined as $s:W\rightarrow\Omega_1$, in such a way that

$$P(s(j)=I_{1,i})=P(I_{2,j}=I_{1,i})=p_{S,i}=\frac{F_i}{\sum_{i=1}^{r}F_i}, i=1,...,r, \forall j=1,...,n, \qquad (3)$$

where the probability that individual i in the original population is copied in position j in the intermediate population, $p_{S,i}$, is defined as the quotient of the fitness value for individual i and the sum of all fitness values for the r individuals in the original population. The intermediate population is obtained by randomly generating the results of n multinomial experiments of size r with probabilities $p_{S,1},...,p_{S,r}$.

2.1.2. Transition matrix

In the second phase, an element of heterogeneity should be incorporated to allow an individual whose utility level is not so high to survive. The crossover and mutation operators are conventionally responsible to do it because they allow the characteristics that identify the selected individual in the first phase to be modified. These operators are the elements of a simple genetic algorithm, but, as Hernández-López and Cáceres-Hernández (2007) pointed out, the transformations in the population that they produce are completely random and not in any way guided by economic statements. Nevertheless, in known social settings, it is possible that information exists that suggests a greater likelihood of some transformations as opposed to others. A good strategy of introducing this information may be to maintain the selection operator only to determine the intermediate population. Once the copies have been selected, a transition matrix is defined in this paper to assign specific probabilities to each one of the transformations. That means to say, in a second phase, each one of the structures copied in the intermediate population remains on the final population on the course $t+1$ or is transformed into another structure according to the probabilities in the corresponding row of the transition matrix as indicated.

Suppose that, given the categories of the individual characteristics, m different structures, $\{E_i\}_{i=1,...,m}$, are defined in such a way that the characteristics of any individual copied at the selection phase correspond to one of these structures. By means of a square transition matrix M, each structure E_i, $i=1,...,m$, from the intermediate population is assumed to be transformed into another structure E_j from the final population with the probabilities $p_T(i,j)$, $j=1,...,m$, located at the i^{th} row of the matrix M.

The formal definition of the transition matrix is as follows. Let $\Omega_2:\{I_{2,1},...,I_{2,n}\}$ be a set of n individuals from the resultant intermediate population from the copies,

E a set of m structures in which each individual in Ω_2 can be transformed, and $Q:\{1,...,n\}$ a set of n positions where individuals from the final population can be placed. This final population $\Omega_3:\{I_{3,1},...,I_{3,n}\}$ is obtained using the transition matrix operator $tm(q)=I_{3,q}$ defined as $tm(q):Q\rightarrow E$, such that

$$P\big(tm(q)=E_j\big)=P\big(I_{3,q}=E_j\big)=p_T(q,j),\ \mathrm{q}=1,...,n,\ j=1,...,m, \qquad (4)$$

is the probability that the individual located at position q in the intermediate population, $I_{2,q}$, is transformed into another individual in the final population whose structure is E_j. If the individual that occupies the position q in the intermediate population has structure E_i, then $p_T(q,j)=p_T(i,j)$, that is, the element in the column j of the row i in the transition matrix. In this way, a multinomial experiment of size m with probabilities $p_T(i,1),...,p_T(i,m)$ can be performed to determine which individual $I_{3,q}$ is transformed from individual $I_{2,q}$.

In this paper, the probabilities in the original transition matrix, A, are calculated in such a way that the shorter the distance between the values of the fitness function corresponding to two different structures, the higher their transition probability. So, if the values of the fitness function corresponding to structures E_i and E_j are equal, then $p_T^A(i,j)=\alpha$, $i,j=1,...,m$; while $p_T^A(i,j)=\beta/\delta_{i,j}$, if the distance between the corresponding values of the fitness function is $\delta_{i,j}$. Of course, $\sum_{j=1}^{m} p_T^B(i,j)=1,\ \forall i=1,...,m$.

2.1.3. Corrected transition matrix

The principles according to which the genetic algorithm is developed may be not able to capture some deviations between the original and the forecasted populations. This being the case, the transition matrix should be corrected. Note that the transition matrix is useful to introduce random changes that are not driven by the same principle as the one that drives the selection operator. Therefore, when the changes observed in the original population are not expected according to the selection operator, this matrix can improve the performance of the genetic algorithm to forecast the composition of a population. In fact, the transition matrix allows the algorithm to take the following points into account. Firstly, the incidence of job opportunities associated with a university degree on the value of the fitness function may need to be adjusted. Secondly, a modification of the selection operator definition could be needed. Thirdly, it may be necessary to introduce relevant factors in the choice of a university degree which are not directly related with job opportunities. For example, the increase in preferences toward civil servant jobs instead of facing up to a life of labor in a private firm, explains some of the changes in demand for university degrees. The question is that, once the algorithm is applied to an original population in such a way that the final population corresponds to a course in which the enrollment on the different degrees is observed, the forecasting performance of the algorithm can be tested, and its design can consequently be adjusted in order to obtain better forecasts of populations on future courses.

The elements of genetic algorithms make it feasible to deduce the expected relative frequency corresponding to each university degree in the final population, and these relative frequencies should be equal to the observed ones. As mentioned, individual characteristics are assumed to generate m different structures, $\{E_i\}_{i=1,...,m}$. Then, the selection probability of structure E_i, $p_S(i)$, depends on the corresponding value of the fitness function and also on the relative frequency of such a structure in the original population. From these probabilities, the selection probabilities corresponding to the degrees in Business Administration, Economics and Management, denoted as $p_S(BA)$, $p_S(ECO)$ and $p_S(MN)$, respectively, can be calculated. Let $p_T^A(i,j)$ be the transition probabilities from structure E_i to structure E_j, and let $p_F(i)$ be the probability that structure E_i is present in the final population. Then,

$$p_F(i) = \sum_{j=1}^{m} p_S(j) p_T^A(j,i), \quad i = 1,...,m, \tag{5}$$

and the final probabilities corresponding to each degree, $p_F(BA)$, $p_F(ECO)$ and $p_F(MN)$, can be easily calculated.

Therefore, when the transformations that the application of transition matrix A involves are observed, the parameters in this matrix can be adjusted to improve the forecasting performance of the genetic algorithm as follows. Let $p_S(D_i)$ be the selection probability corresponding to degree D_i, let $p_T(D_i, D_j)$ be the transition probability from degree D_i to degree D_i, and let $p_F(D_i)$ be the probability that degree D_i is present in the final population. Then,

$$\begin{aligned} &p_F(BA) = \\ &p_S(BA)\, p_T(BA,BA) + p_S(ECO)\, p_T(ECO,BA) + p_S(MN)\, p_T(MN,BA) \end{aligned} \tag{6}$$

$$\begin{aligned} &p_F(ECO) = \\ &p_S(BA)\, p_T(BA,ECO) + p_S(ECO)\, p_T(ECO,ECO) + p_S(MN)\, p_T(MN,ECO) \end{aligned} \tag{7}$$

$$\begin{aligned} &p_F(MN) = \\ &p_S(BA)\, p_T(BA,MN) + p_S(ECO)\, p_T(ECO,MN) + p_S(MN)\, p_T(MN,MN) \end{aligned} \tag{8}$$

Probabilities $p_S(BA)$, $p_S(ECO)$ and $p_S(MN)$ depend on the selection operator and the composition of the original population. In order to obtain an unbiased forecast, the final probabilities $p_F(BA)$, $p_F(ECO)$ and $p_F(MN)$ should be equal to the relative frequencies corresponding to new students enrolled on these degrees on course $t+1$. However, the elements of the transition matrix

$$\begin{pmatrix} p_T(BA,BA) & p_T(BA,ECO) & p_T(BA,MN) \\ p_T(ECO,BA) & p_T(ECO,ECO) & p_T(ECO,MN) \\ p_T(EMP,BA) & p_T(MN,ECO) & p_T(MN,MN) \end{pmatrix} \tag{9}$$

cannot be derived from equations (6) to (8). It is clear that

$$p_T(BA,BA)+p_T(BA,ECO)+p_T(BA,MN)=1, \tag{10}$$

$$p_T(ECO,BA)+p_T(ECO,ECO)+p_T(ECO,MN)=1 \tag{11}$$

$$p_T(MN,BA)+p_T(MN,ECO)+p_T(MN,MN)=1, \tag{12}$$

but additional constraints are required. In this sense, the decision is made to fix the non-transition probabilities

$$p_T(BA,BA)=k_A,\ p_T(ECO,ECO)=k_B,\ p_T(MN,MN)=k_C, \tag{13}$$

and another element of the transition matrix between degrees. Finally, in order to match probability $p_F(D_i)$ with the relative frequency corresponding to degree D_i observed in the final population, the following constraints in the transition matrix between structures are also imposed:

$$\begin{aligned} &\sum_{j\in BA} p_T(j,i)=p_T(BA,BA) \\ &\sum_{j\in ECO} p_T(j,i)=p_T(BA,ECO), \\ &\sum_{j\in MN} p_T(j,i)=p_T(BA,MN) \end{aligned} \tag{14}$$

for structures i in which the degree is Business Administration,

$$\begin{aligned} &\sum_{j\in BA} p_T(j,i)=p_T(ECO,BA) \\ &\sum_{j\in ECO} p_T(j,i)=p_T(ECO,ECO) \\ &\sum_{j\in MN} p_T(j,i)=p_T(ECO,MN) \end{aligned} \tag{15}$$

for structures i in which the degree is Economics, and

$$\sum_{j \in BA} p_T(j,i) = p_T(MN, BA)$$
$$\sum_{j \in ECO} p_T(j,i) = p_T(MN, ECO) \qquad (16)$$
$$\sum_{j \in MN} p_T(j,i) = p_T(MN, MN)$$

for structures i in which the degree is Management. Furthermore, the partial sums corresponding to the transition probabilities between degrees are distributed between elements $p_T(j,i)$ corresponding to each one of them in such a way that the values of parameters $p_T(j,i)$ in the corrected transition matrix between structures are proportional to the original values of these probabilities in transition matrix A. So, a new transition matrix between structures, B (correction of matrix A), is obtained. Consequently, two different genetic algorithms are also designed.

2.2. Data sources

Before applying the designed genetic algorithm, two sources of data need to be taken into account. First, a population of graduates is needed to estimate the probabilities that one of them remains unemployed, or gets a temporary job or, finally, gets a permanent job; and, from these probabilities, a fitness function is defined that measures the ability of a graduate to access the labor market. Second, populations of new students enrolled on specific university degrees on some courses constitute the original populations to which the genetic algorithm is applied to forecast the composition of these populations for the following courses.

2.2.1. Graduate population and estimates of job opportunities

Students are assumed to assign expectancies of getting a type of job to different university degrees which are obtained from the observation of the labor situation of graduates from past courses. *Observatorio Permanente para el Seguimiento de la Inserción Laboral* of the University of La Laguna provides information about the labor activity of graduates in Business Administration, Economics or Management in 2004 and 2005 during two years after finishing their degree. Note that the conclusions about job opportunities of graduates depend on the length of the time period after graduation which is analyzed. The access to the labor market may be delayed as a consequence of the students decision making about the continuation of their education, but the period analyzed should not be excessively long because further education and, above all, learning derived from the labor experience weaken the influence of the degree.

According to Biggeri, Bini and Grilli (2001) and Salas (2007), those graduates who have previous working experience are more likely to obtain a job sooner. To avoid this circumstance distorting the analysis of the effect of the degree on job opportunities, data corresponding to graduates with labor experience before graduation are eliminated from the original population. As results in Salas (2007) show, a key factor which explains the time needed to obtain a first employment for higher education graduates is the job search effort. So, graduates who opted for self-employment and those who decided not to get a job are also eliminated. Finally, graduates who only got temporary jobs in such a way that their labor experience is shorter than three months, are included as unemployed.

For each individual in the population of graduates, the following characteristics are available: Y (labor status), with categories 1 (unemployed), 2 (temporary job) and 3 (permanent job); $X(D)$ (degree), with categories 1 (Business Administration), 2 (Economics) and 3 (Management); $X(A)$ (age), defined as the difference between the year of graduation and birthdate with categories 1 (younger than 25 years old), 2 (25 to 26 years old) and 3 (older than 26 years old); $X(G)$ (gender), with categories 1 (male) and 2 (female). Then, probabilities of getting a temporary job or a permanent job, and the probability of remaining unemployed two years after graduation are obtained by estimating a multinomial logit model in which the explanatory variables are dichotomic variables corresponding to each one of the categories of degree, age and gender. The results of estimating such a model, available from authors upon request, show that temporary jobs are more likely than permanent jobs, but the most likely labor status of recent graduates is unemployed.

The age of graduation and, above all, the university degree are important factors for the access of the graduates to the labor market. In fact, although graduates on different degrees are able to develop the same job, and a specific degree can provide the ability to develop jobs not conventionally related to such a program study, the university degree is an usual explanatory variable for job opportunities (Rodríguez and Gutiérrez 2007). The effect of gender is not so significant. In this sense, Olave and Salvador (2006) indicate that there are gender differences in the access to labor markets in the long term, but not in the short term. However, Rodríguez and Gutiérrez (2007) find that a male is more likely to get a job. On the other hand, preliminary estimates were obtained including the education level of parents and their labor activity as explanatory variables, but they were not statistically significant.

According to estimated probabilities, the degree in Business Administration provides the best job opportunities. The average probability that a graduate on this degree gets a permanent job is twice the same probability for graduates in Economics, which is the degree with the highest probability of remaining unemployed. To evaluate effects of changes in categories of individual characteristics, quotients of odds-ratios are also calculated, but not shown to save space. Such quotients reveal that, for graduates in Business Administration, the ratio between the probability of getting a permanent job and the probability of remaining unemployed is more than 10 percentual points higher than in the case of graduates in Management, and almost two and a half times the value of such a ratio for graduates in Economics. On the other hand, the effect of age on job opportunities is non-linear. The access to the labor market is more likely when the age of the graduate is 25 to 26 years old. These graduates have a clearly higher probability of getting a permanent job and a lower probability of remaining unemployed than the remaining graduates. This circumstance is clear from the quotient of ratios between these probabilities for graduates with different ages. With regards to gender, there are no remarkable differences in the probability of getting a permanent job. However, there is a slightly stronger trend towards temporary jobs for females, whereas the probability of remaining unemployed is higher for males.

Anyway, the expectancy of accessing the labor market should not only be measured in terms of the previous probabilities, but the length of studies should also be taken into account. In this sense, the degree in Management provides the possibility of an earlier access to the labor market, and, therefore, the job opportunities of graduates in Management are actually better than suggested by the logit model estimates. In this sense, in accordance to Equations (1) and (2), Table 1 shows the values of the fitness function corresponding to each one of the 18 different structures which can be defined from categories of individual characteristics, that is, three degrees, three age segments and two genders.

Table 1. Estimates of the fitness function for the individual structures

Structure	Graduates	$P(Y=1)$	$P(Y=2)$	$P(Y=3)$	Fitness
1 Business Adm./Younger than 25/Male	11	0.4678	0.2886	0.2436	3.3020
2 Business Adm./Younger than 25/Female	36	0.3757	0.3720	0.2524	3.6394
3 Business Adm./25 or 26 years old/Male	3	0.3732	0.2808	0.3460	4.3025
4 Business Adm./25 or 26 years old/Female	19	0.2938	0.3549	0.3513	4.5780
5 Business Adm./Older than 26/Male	15	0.5299	0.2438	0.2263	2.9942
6 Business Adm./Older than 26/Female	13	0.4368	0.3226	0.2406	3.3736
7 Economics/Younger than 25/Male	7	0.5816	0.2919	0.1265	2.1405
8 Economics/Younger than 25/Female	13	0.4794	0.3862	0.1345	2.5031
9 Economics/25 or 26 years old/Male	11	0.5002	0.3062	0.1936	2.8550
10 Economics/25 or 26 years old/Female	15	0.4029	0.3959	0.2012	3.1997
11 Economics/Older than 26/Male	17	0.6441	0.2411	0.1148	1.8716
12 Economics/Older than 26/Female	26	0.5462	0.3281	0.1256	2.2407
13 Management/Younger than 25/Male	57	0.4455	0.3458	0.2086	4.1652
14 Management/Younger than 25/Female	93	0.3509	0.4372	0.2119	4.5745
15 Management/25 or 26 years old/Male	32	0.3596	0.3405	0.2998	5.3599
16 Management/25 or 26 years old/Female	62	0.2781	0.4228	0.2991	5.6790
17 Management/Older than 26/Male	38	0.5095	0.2949	0.1956	3.7879
18 Management/Older than 26/Female	53	0.4125	0.3833	0.2043	4.2566

2.2.2. Populations of new students enrolled on university degrees

Once the fitness function is defined from the estimates of probabilities that graduates in 2004 and 2005 access the labor market, the genetic algorithm is applied to forecast the transformations in the original populations of new students enrolled on courses 2005/2006 to 2008/2009.

The characteristics of new students in original populations are also defined in terms of the individual characteristics that, according to the logit model, are associated to job opportunities, that is, degree, gender and age. However, in the definition of the fitness function one of the explanatory variables is graduation age. Given that the original populations are composed of new students enrolled on the first course of the degree, the age categories should be corrected. The student expectancies to finish a degree can be assumed to depend on the planned lengths of studies, and, therefore, the definition of the fitness function does not need to be modified. However, the unknown age of graduation is a characteristic to be estimated from the age of admission. In this sense, the actual length of the studies seems to be a more appropriate parameter than the official planned length. Due to that, the age of graduation could be evaluated from the actual average length of the studies, but this procedure implies that a graduate in Economics could not be younger than 25 years old. Furthermore, the average length of studies may be overestimated as a consequence of calculations made including students who got a job before finishing their degree, and they usually complete their studies later than students who try to get a job after finishing the degree. Given the nature of the genetic algorithm, the decision is made to modulate the effects of the actual lengths of the studies. Then, the age categories are finally defined as 1, 2 or 3, if the admission age plus 6 (in the case of the Business Administration degree), 7 (in the case of the Economics degree) or 5 years (in the case of the Management degree) is less than 25 years old, equal to 25 to 26 years old or higher than 26 years old.

On the other hand, the demand for the degrees in question is not directly observed due to the access to some of them being limited by quotas. So, the number of newly enrolled students does not reveal the actual demand. During the period analysed in this paper, the access to the degree in Economics is not limited, but there are quotas to the degrees in Business Administration and Management. Then, the demand for the degrees is identified as the number of pre-enrollments in June for degrees in Business Administration and Management, whereas in the case of the degree in Economics the demand is defined as the number of students who passed the admission tests to University in June and enrolled on the degree for the following course. Note that the total number of students enrolled on Economics overestimates the actual demand for this degree, due to a significant group of these students corresponding to those who are not able to access the degree in Business Administration.

Moreover, a forecast of the population of new students enrolled on course 2006/2007 is going to be obtained from the estimates of probability that graduates in 2004 or 2005 access the labor market two years after their graduation. This forecast is only useful in evaluating the forecasting performance of the genetic algorithm and to improve its ability to forecast future populations. Given that data about the number of new students enrolled on courses 2005/2006 to 2008/2009 are available from the *Gabinete de Análisis y Planificación* of the University of La Laguna, these populations are the original populations to which the genetic algorithm is applied to forecast the composition of populations in the following course. The final population obtained as result of executing the algorithm could be assumed as being the original population in forecasting the demand for the following course, but, as a consequence of the quotas mentioned, the number of pre-enrollments does not match the number of new students enrolled. Therefore, the populations composed of new students enrolled on the first course of the degrees under analysis are chosen to be the original populations of the genetic algorithm. The relative weights of degrees, age segments and genders in populations of new students enrolled on the first course of the degrees on courses 2005/2006 to 2008/2009 are available from authors upon request, whereas Table 2 shows the demand observed for the degrees in Business Administration, Economics and Management on courses 2006/2007 to 2008/2009.

Table 2. Composition of the demand for degrees

	Degree (number of students)			Degree (%)		
Course	BA	ECO	MN	BA	ECO	MN
06/07	207	121	420	0.277	0.161	0.561
07/08	211	395	123	0.2894	0.1687	0.5418
08/09	192	357	116	0.2887	0.1744	0.5368

3. Results and discussion

In populations composed of new students enrolled on the first course of a degree, each individual is identified by one of the 18 different structures defined from the degrees, segments of age, and genders analyzed. So, a value of the fitness function can be assigned to each individual in these populations according to the estimates of the logit model, and the results of applying the selection operator depend on these values. The intermediate population obtained from the selection is transformed by applying the transition matrix A, defined by assuming that the non-transition probabilities are equal to 0.70.

By applying the selection operator, an intermediate population is generated, which is composed of 1000 individuals to facilitate the calculation of the relative weights of the different degrees, given that this is also the number of individuals in the final population. When the genetic algorithm is applied to the original populations of new students enrolled on courses 2005/2006 to 2007/2008 and final populations of new students enrolled on courses 2006/2007 to 2008/2009 are forecasted, the corresponding corrected transition matrix B is defined in such a way that the expected relative weights of the degrees in forecasted populations are equal to the relative weights of the degrees in the observed demand. From the corrections applied to generate these forecasts, a specific corrected transition matrix B is calculated. Finally, this matrix is used to transform the intermediate population resulting from applying the selection operator to the original population composed of new students enrolled on course 2008/2009. Thus, a forecast of the new student population enrolled on course 2009/2010 is obtained.

3.1. Forecasts of populations on courses 2006/2007 to 2008/2009

The genetic algorithm provides forecasts of the relative weight of structures $\{E_i\}_{i=1,...18}$ in the population of new students enrolled on courses 2006/2007 to 2008/2009, which are obtained from the original populations composed of new students enrolled on courses 2005/2006 to 2007/2008. For each one of these courses, the probabilities of selecting structure E_i in the selection phase, $p_S(i)$, and the probabilities that structure E_i is present in the final population, $p_F^A(i)$, is obtained when transition matrix A is applied. The final probabilities corresponding to each degree are shown in Table 3. These probabilities define the expected relative weights of structures and degrees in the intermediate and final populations.

Table 3. Genetic algorithms and forecasted population on courses 2006/2007 to 2008/2009

	2006/2007			2007/2008			2008/2009		
Degree	BA	ECO	MN	BA	ECO	MN	BA	ECO	MN
Selection and predicted final probabilities									
$p_S(i)$	0.2407	0.1673	0.5920	0.2323	0.1721	0.5955	0.2440	0.1656	0.5904
$p_F^A(i)$	0.3276	0.1640	0.5084	0.3215	0.1689	0.5095	0.3305	0.1651	0.5044
Forecasting performance of genetic algorithms									
Demand	277	162	561	289	169	542	289	174	537
GA. A	341	184	475	325	172	503	317	147	536
Test[a]	$\chi_2^2 = 31.3*$			$\chi_2^2 = 7.217**$			$\chi_2^2 = 7.086**$		
GA B	261	181	558	269	161	570	302	150	548
Test[a]	$\chi_2^2 = 3.204$			$\chi_2^2 = 3.26$			$\chi_2^2 = 4.266$		

[a] Statistically significant at 1% level (*), at 5% level (**), or at 10% level (***).

In view of the relative weight of the three degrees in the original populations of new students and the observed demand for these degrees on the following courses, the changes in the composition of populations are in accordance with the expected forecasts obtained by applying genetic algorithms. Note that the algorithm should lead to the reduction of the demand for the degree in Economics, which has the lowest values of the fitness function, whereas the increases in the relative weights of the degrees in Business Administration and Management are also predictable. Nonetheless, the estimates of selection probabilities suggest that the selection operator leads to undervalue the demand for Business Administration and to overestimate the demand for Management. On the other hand, the final probabilities show the effect of the transition matrices. The transformation of the original population as a consequence of the conjoint action of selection operator and transition matrix is not enough to adjust the composition of the demand on the following course. The demand for Business Administration is overestimated, the demand for Management is undervalued, and the demand for Economics is well adjusted on courses 2006/2007 and 2008, but undervalued on course 2008/2009.

Once the transformations derived from the application of transition matrix A are observed, the algorithm should be corrected to improve its forecasting performance. Following the procedure explained in methodological section, the correction of the transition matrix between structures is deduced from the previous calculation of the parameters of a transition matrix between degrees, as defined in

(9). Taking into account that the higher the non-transition probabilities, the weaker the distortion on the effect of the selection operator, and once several alternatives are tested, the decision is made to apply the following constraints: $p_T(BA, BA) = 0.6$, $p_T(ECO, ECO) = 0.3$, and $p_T(MN, MN) = 0.7$. In addition, it is also assumed that $p_T(MN, ECO) = 0.1$. That is, the transition probability from the most popular degree to the least attractive one is fixed at a low level. From these constraints, and according to the selection parameters, the parameters of the transition matrices between degrees which are applied to forecast the populations on courses 2006/2007, 2007/2008 and 2008/2009 are obtained.

Finally, in order that the probabilities $p_F(i)$ match the relative frequencies corresponding to the degree i observed in the final population, the following constraints in the transition matrix between structures are also imposed:

$$\begin{aligned}
&\sum_{j=1}^{6} p_T(j,i) = 0.6,\ i = 1,...,6\\
&\sum_{j=7}^{12} p_T(j,i) = p_T(BA, ECO),\ i = 1,...,6\\
&\sum_{j=7}^{12} p_T(j,i) = p_T(BA, MN), i = 1,...,6
\end{aligned} \tag{17}$$

for structures i in which the degree is Business Administration,

$$\begin{aligned}
&\sum_{j=1}^{6} p_T(j,i) = p_T(ECO, BA),\ i = 7,...,12\\
&\sum_{j=7}^{12} p_T(j,i) = 0.3,\ i = 7,...,12\\
&\sum_{j=7}^{12} p_T(j,i) = p_T(ECO, MN),\ i = 7,...,12
\end{aligned} \tag{18}$$

for structures i in which the degree is Economics, and

$$\begin{aligned}
&\sum_{j=1}^{6} p_T(j,i) = 0.20,\ i = 13,...,18\\
&\sum_{j=7}^{12} p_T(j,i) = 0.10,\ i = 13,...,18\\
&\sum_{j=7}^{12} p_T(j,i) = 0.70,\ i = 13,...,18
\end{aligned} \tag{19}$$

for structures i in which the degree is Management. The partial sums corresponding to the transition probabilities between degrees are distributed between terms $p_T(j,i)$ in the corrected transition matrix between structures as indicated in methodological section. The new transition matrix B provides unbiased forecasts of the demand for the degrees. This matrix is not shown in the paper, but it is available upon request. Table 3 also shows the results of goodness of fit tests for the forecasted populations obtained by applying each one of the two different genetic algorithms corresponding to the application of matrices A and B.

3.2. Forecast of population on course 2009/2010

In the following paragraphs, the results of applying the algorithm to the original population composed of the new students enrolled on the course 2008/2009 are set out. Table 4 shows the selection probabilities, $p_S(i)$, and, also, the final probabilities, $p_F^A(i)$ and $p_F^B(i)$, corresponding to the different degrees, when matrices A or B are introduced into the genetic algorithm which is applied to forecast the population on course 2009/2010.

Table 4. Genetic algorithms and forecasted population on course 2009/2010

Degree	Selection and predicted final probabilities			Composition of the forecasted population	
	$p_S(i)$	$p_F^A(i)$	$p_F^B(i)$	GA. Matrix A	GA. Matrix B
Business Administration	0.2381	0.3291	0.2952	32.2%	29.0%
Economics	0.1304	0.1381	0.1729	14.5%	16.5%
Management	0.6316	0.5328	0.5319	53.3%	54.5%

The forecasting performance of the genetic algorithm on courses 2006/2007 to 2007/2008 suggests that the selection operator and transition matrix A are not expected to provide adjusted forecasts of the relative weights of the three degrees on course 2009/2010. Therefore, the parameters of these transition matrices should be corrected to obtain a new matrix B, which is defined from the corrected transition matrix in the case of the algorithms applied to forecast the demand on courses 2006/2007 to 2008/2009. In view of the similarities between the three matrices, the decision is made to calculate the parameter value at a specific position in the corrected transition matrix between degrees as the result of a linear adjustment to the calculated values at the same positions in these transition matrices. So, the following transition matrix between degrees is obtained.

$$\begin{pmatrix} p_T(BA,BA) & p_T(BA,ECO) & p_T(BA,MN) \\ p_T(ECO,BA) & p_T(ECO,ECO) & p_T(ECO,MN) \\ p_T(EMP,BA) & p_T(MN,ECO) & p_T(MN,MN) \end{pmatrix} = \begin{pmatrix} 0.60 & 0.30 & 0.10 \\ 0.20 & 0.30 & 0.50 \\ 0.20 & 0.10 & 0.70 \end{pmatrix} \quad (20)$$

Then, a corrected matrix between structures, B, is obtained by imposing the following constraints:

$$\sum_{j=1}^{6} p_T(j,i) = 0.60, \sum_{j=7}^{12} p_T(j,i) = 0.30, \sum_{j=7}^{12} p_T(j,i) = 0.10, i = 1,\ldots,6, \quad (21)$$

$$\sum_{j=1}^{6} p_T(j,i) = 0.20, \sum_{j=7}^{12} p_T(j,i) = 0.30, \sum_{j=7}^{12} p_T(j,i) = 0.50, i = 7,...,12\,, \tag{22}$$

$$\sum_{j=1}^{6} p_T(j,i) = 0.20, \sum_{j=7}^{12} p_T(j,i) = 0.10, \sum_{j=7}^{12} p_T(j,i) = 0.70, i = 13,...,18\,, \tag{23}$$

and distributing the values of partial sums as indicated in methodological section. This matrix is not shown to save space, but it is available from authors upon request. Table 4 shows the composition of the population of new students enrolled on course 2009/2010 generated by an application of the different genetic algorithms to the original population of new students enrolled on course 2008/2009. From the corrected matrix B and the selection probabilities, the final probabilities that determine the relative weight corresponding to each degree in the forecasted population are also shown in Table 4. Note that according to the results of the application of the genetic algorithms with the corrected transition matrix, almost 30% of the new students enrolled on course 2009/2010 are expected to choose the Business Administration degree, whereas around 17% will opt for Economics and more than a half will try to study Management.

4. Conclusions

The genetic algorithm proposed in this paper forecasts the relative weight corresponding to different degrees in the demand for higher education of a group of university students as a whole. Note that the population forecasted each time the genetic algorithm is executed may be different, but the forecast of the relative weight corresponding to a degree is unlikely to be far from its corresponding final probability, that is, its expected relative weight.

Of course, generalization of this methodology to a new context is an empirical matter whose evaluation may be made by the reader. The implementation of new program studies at an early date in European universities would require a new design of the algorithm in order to get good forecasting performance. Nevertheless, the proposed methodology is flexible enough to adapt to more general settings. For example, specific genetic algorithms could be designed to forecast the demand for different studies supplied by a university, or, the demand for different sets of degrees framed in scientific fields. This being the case, the main obstacles to overcome are that different probabilities to get a job may be associated to different degrees framed into the same scientific field and, above all, that vocations are more difficult to be left aside as a factor to explain the choice of an university degree.

Anyway, the specific results obtained in this paper are not its main contribution to the literature, but the proposal of a methodology able to forecast the relative weight of a degree in the demand for higher education as a function of the probabilities of accessing the labor market associated with the degree. This methodological proposal is shown to be able to provide the type of forecast which is very useful in policy making decisions in the recent process of building Higher Education Areas.

Acknowledgement
This research would not have been possible without the assistance of the Gabinete de Análisis y Planificación and the Observatorio Permanente para el Seguimiento de la Inserción Laboral of the University of La Laguna.

REFERENCES

[1] **Aguilar, M.I., and García, D. (2008),** ***Desajuste educativo y salarios en España: nueva evidencia con datos de panel***. *Estadística Española*, 50, 393-426;
[2] **Barceinas, F., Oliver, J., Raymond, J.L., and Roig, J.L. (2000),** ***Los rendimientos de la educación y la inserción laboral en España***. *Papeles de Economía Española*, 86, 128-148;
[3] **Becker, G.S. (1993),** ***Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education***. Chicago: Chicago University Press;
[4] **Bellas, M.L. (2001),** ***Investment in Higher Education: Do Labor Market Opportunities Differ by Age of Recent College Graduates***. *Research in Higher Education*, 42, 1-25;
[5] **Biggeri, L., Bini, M. and Grilli, L. (2001),** ***The Transition from University to Work: A Multilevel Approach to the Analysis of the Time to Obtain the First Job*.** *Journal of the Royal Statistical Society Series A-Statistics in Society*, 164, 293-305;
[6] **Budría, S. and Telhado-Pereira, P. (2011),** ***Educational Qualifications and Wage Inequality: Evidence for Europe***. *Revista de Economía Aplicada*, 19, 5-34;
[7] **De Paola, M. and Gioia, F. (2012),** ***Risk Aversion and Field of Study Choice: The Role of Individual Ability***. *Bulletin of Economic Research*, 64, 193-209;
[8] **Dolado, J.J., Felgueroso, F. and Jimeno, J.F. (2000),** ***La inserción laboral de los titulados universitarios en España***. *Papeles de Economía Española*, 86, 78-97;
[9] **Fernández, C. (2006),** ***The Role of Education Vis-a-vis Job Experience in Explaining the Transitions to Employment in the Spanish Youth Labour Market*.** *Spanish Economic Review*, 8, 161-187.;
[10] **García-Montalvo, J. (2005),** ***La inserció laboral dels universitaris: la question de la sobrequalifiació i el desajust formatiu*.** *Nota d'Economia*, 1, 99-119;
[11] **Goldberg, D.E. (1989),** ***Genetic Algorithms in Search Optimization and Machine Learning***. New York: *Addison Wesley*;
[12] **Grilli, L. and Mealli, F. (2008),** ***Nonparametric Bounds on the Causal Effect of University Studies on Job Opportunities Using Principal Stratification***. *Journal of Educational and Behavioral Statistics*, 33, 111-130;
[13] **Hartog, J. and Díaz-Serrano, L. (2007),** ***Earning Risk and Demand for Higher Education: A Cross Section Test for Spain***. *Journal of Applied Economics*, 10, 1-28;
[14] **Hernández-López, M. and Cáceres-Hernández, J.J. (2007),** ***Forecasting Tourists' Characteristics by a Genetic Algorithm with a Transition Matrix***. *Tourism Management*, 28, 290-297;

[15**] Holland, J.H. (1975),** ***Adaptation in Natural and Artificial Systems***. Michigan, Ann Arbor: *The University of Michigan Press*;
[16] **Kucel, A. and Vilalta-Bufí, M. (2013),** ***Job Satisfaction of University Graduates***. *Revista de Economía Aplicada*, 21, 29-55;
[17] **Lassibille, G., Navarro, L., Aguilar, I. and Sánchez, C. (2001),** ***Youth Transition from School to Work in Spain***. *Economics of Education Review*, 20, 134-149;
[18] **Latiesa, M. (1989),** ***Demanda de educación superior: evaluación y condicionamientos de los estudiantes en la elección de carrera***. *Revista Española de Investigaciones Sociológicas*, 46, 101-139;
[19] **Mora, J.G. (2008),** ***El éxito laboral de los jóvenes graduados universitarios europeos***. *Revista de Educación*, número extraordinario, 41-58;
[20] **Navarro, G. and Casero, A. (2012),** ***Analysis of Gender Differences in Degree Choice***. *ESE-Estudios sobre Educación*, 22, 115-132.;
[21] **Nieto, S. and Ramos, R. (2013),** ***Non-Formal Education, over Education and Wages***. *Revista de Economía Aplicada*, 21, 5-28;
[22] **Olave, M.P. and Salvador, M. (2006),** ***Impacto de la formación ocupacional en el primer período de desempleo de egresados universitarios en Aragón. Una aproximación semiparamétrica bayesiana***. *Estadística Española*, 48, 101-134;
[23] **Peró, M., Soriano, P.P., Capilla, R., Guardia i Olmos, J. and Hervás, A. (2015),** ***Questionnaire for the Assesment of Factors Related to University Degree Choice in Spanish Public System: A Psychometric Study***. *Computers in Human Behavior*, 47, 128-138;
[24] **Rodríguez, C. and Gutiérrez, J. (2007),** ***Empleo de modelos de regresión logística binomial para el estudio de variables determinantes en la inserción laboral de egresados universitarios***. *Investigación y Postgrado*, 22, 109-144;
[25] **Sáez, F. and Rey, R. (2000),** ***La inserción laboral de los universitarios***. *Papeles de Economía Española*, 86, 99-110;
[26] **Salas, M. (1996),** ***La regresión logística. Una aplicación a la demanda de estudios universitarios***. *Estadística Española*, 38, 193-217;
[27] **Salas, M. (2004),** ***La relación educación-economía: un estudio del desajuste educativo de los titulados universitarios***. *Revista de Educación*, 334, 259-278;
[28] **Salas, M. (2005),** ***La demanda de educación superior: un análisis microeconómico con datos de corte transversal***. *Revista de Educación*, 337, 327-349;
[29] **Salas, M. (2007),** ***The Transition from Higher Education to Employment in Europe: The Analysis of the Time to Obtain the First Job.*** *Higher Education*, 54, 333-360.
[30] ***Skatova, A. and Ferguson, E. (2014), Why Do Different People Choose Different University Degrees? Motivation and the Choice of Degree***. *Frontiers in Psychology*, 5(1244), 1-15;
[31] **Tribe, K. (2003),** ***Demand for Higher Education and the Supply of Graduates.*** *European Educational Research Journal*, 2, 463-471.