Revista de Investigación en Educación, nº 11 (1), 2013, pp. 7-22 ISSN: 1697-5200
http://webs.uvigo.es/reined/ eISSN: 2172-3427

ARTÍCULO ORIGINAL - COLABORACIÓN ESPECIAL

Determination of the teachers' workload in the framework of the EHEA

Antonio García Pino
agpino@uvigo.es
Universidade de Vigo

ABSTRACT. This paper describes a quantitative model to determine the required faculty human resources for developing higher education degree. The parameters that intervene in the model are thus presented and discussed. There are two kinds of parameters: a) academic features such as duration of studies, number of students entering each year, graduation, dropout and efficiency rates, etc. b) parameters derived from regulations. The model is based on the stability of some of the academic parameters in a long period. A mathematical development leads to a practical formulation to predict the number of students enrolled, the number of credits inscribed and the number of faculty hours required. Some approximations on the results are also suggested in order to achieve very simple formulas which are useful for the bodies of government of the universities.

KEY WORDS. Faculty Members, Graduate, Dropout, Curriculum, Human Resources

Determinación de la carga de trabajo de los profesores en el marco del Espacio Europeo de Educación Superior

RESUMEN. Se presenta un modelo cuantitativo para determinar las necesidades de profesorado en educación superior. Se presentan y discuten los parámetros que intervienen en el modelo, que se clasifican en dos tipos: a) indicadores académicos como duración de los estudios, número anual de estudiantes de nuevo ingreso, tasas de graduación, abandono y eficiencia, etc. b) parámetros que se derivan de las normativas. El modelo se basa en la estabilidad de ciertos parámetros académicos a lo largo de periodos de varios años. Se desarrolla una formulación matemática que permite predecir el número de estudiantes, de créditos matriculados y de horas lectivas necesarias de profesorado. Se introducen algunas aproximaciones en los resultados que conducen a la propuesta de fórmulas muy sencillas y útiles para los órganos de gobierno de las universidades.

PALABRAS CLAVE. Profesores, Graduados, Abandono Escolar, Currículum, Recursos Humanos

1. INTRODUCTION

The European Credit Transfer System, ECTS (European Commission, 2004) allows quantifying the student effort in the European

Fecha de recepción 14/12/2012 · Fecha de aceptación 31/01/2012
Dirección de contacto:
Antonio García Pino
Escola de Enxeñaría de Telecomunicación
Campus As Lagoas-Marcosende. Universidade de Vigo

Higher Education Area, EHEA (Tovar, 2007). By definition, each ECTS involves between 25 and 30 hours of work by the students. These hours include both those developed in classroom or laboratory under the direct intervention of teachers, and the autonomous activity by the student. In annual average values, students take 60 credits a year, representing between 1,500 and 1,800 hours of academic activity. However, there is not a direct measure that relates the credits taken by students with the effort that represents in hours of work by the teaching staff. The workload of teachers is a crucial element from the economic point of view because an important part of higher education costs is dedicated to maintain its human resources.

The number of teachers which is necessary in the long term for developing high education studies depends on different parameters that can be classified as follows:

- Academic features such as the duration of the program, the number of students entering every year, graduation, dropout and efficiency rates, etc.

- Parameters derived from regulations. Some of them allow intervening on the quality of higher education, for instance the average size of the groups in the different activities, and the percentage of contact hours for students in the ECTS. Other parameters can derive from education and labor law, such as the number of hours available annually by the faculty for the exercise of the teaching activities.

A mathematical and quantitative model is presented in this paper in order to predict the workload of educators and, consequently, the dimension of the faculty staff, given a set of simple parameters which can be considered stable for a period of several years. The model will be illustrated with some examples and it can be considered a useful tool for the bodies of government of the universities with responsibility in establishing adequate dimensions for the educational human resources. Since the input parameters are supposed to evolve slowly, the model allows for planning the necessary evolution of these human resources.

A similar objective model was previously presented (Garcia Pino, 2012), but it had certain restrictions on the parameters that were handled. In the formulation presented here, the model is widespread. Besides, a parametric study has been developed which demonstrates the robustness of the model.

2 ACADEMIC PARAMETERS OF THE MODEL

The academic parameters are related to the transit of students along the different levels or years of the studies that constitute the curriculum of the university degree. The mathematical model which is proposed in this work assumes that these parameters are stable in a long term period of several years. Practically, this means that the value of the parameters that must incorporated in the model should be obtained statistically as the mean values computed in a fixed long term period. Some of the parameters act as input parameters obtained directly from such statistics while others are established by the model with the formulation that is described in the following lines.

2.1. Number of students entering studies (N)

The number of students entering the first year of the degree for the first time is denoted by N and is considered an input parameter obtained from the historical records of the degree. The students entering the program in levels/years higher than the first, coming from other program of studies, are not included in this parameter. This number is the most important one in the process because it is linked to the existing social demand of the degree and is directly related to the teachers' workload. If a percentage of the students have part-time status, the parameter N must be computed in this model as the equivalent number of full time-students.

2.2. Number of years (N_y) and credits (N_c) of the studies

A study program or degree consists of a number of ECTS credits (N_c) that must be overcome by the students to achieve the degree. Those credits are distributed along a number of

N_y years or courses. Usually an amount of 60 ECTS per year is assumed for all the programs. When part of the credits of the curriculum does not require regular classes with the presence of teachers, that part can be considered excluded from the parameter N_c. Besides, an effective N_c can be also computed for this model in order to consider the input of students in levels/years higher than the first one. The mathematical model will be illustrated here with an example for the undergraduate degree called "Grado" in Spanish regulations (Spanish regulation RD 1393, 2007), which consists of a 4-year program with 240 ECTS.

2.3. Graduation rate (T_g) and dropout or abandon rate (T_a)

2.3.1 Mathematical model for the dropout rate

According to Spanish regulations (Spanish regulation RD 1393, 2007), the dropout rate (Tinto, 1975; Cabrera, 2006) is defined as the percentage of students (relative to their cohort of new income) that must have obtained the degree in the previous academic year but have not been enrolled or graduated in that academic year or in the previous one. This is a statistical input parameter of the model. A more detailed description of the dropout, year by year, must be considered in order to accomplish an adequate mathematical model allowing predictions of the total number of students enrolled in the studies program at a specific time. The following formulation is proposed: for the first year $T_a^{(1)}$ represents the percentage of students, relative to their cohort of new income, that are only enrolled during the first year and then leave the program without starting a second year. Statistically, this is determined by non-student enrolment during the two subsequent academic years (second and third). The second term $T_a^{(2)}$ represents the percentage of students enrolled only during two academic years (not enrolled during third and fourth). The term $T^{(i)}$ represents the percentage of students enrolled during i years and not enrolled during $i+1$ and $i+2$. With this formulation, the official dropout rate for a degree, T_a, is the sum of the first N_y-*1* terms of the series. The sum of the rest of the terms is denoted here by t_a which represents the percentage of students that leave the studies without graduating after N_y or more years of being enrolled. This is referred here as *late dropout rate*. In general, in the present model, the annual dropping out is assumed to be given by a probability distribution characterized by the point probabilities a_i corresponding to each year as follows:

$$T_a^{(i+1)} = (T_a + t_a)a_i \quad (i \geq 0) \tag{1}$$

where a_i is the probability of abandon in the year i, which is an statistical variable, usually decreasing, with a mean value α. The following relations must be satisfied:

$$\sum_{i=0}^{\infty} a_i = 1 \quad ; \quad \sum_{i=0}^{\infty} i \cdot a_i = \alpha \tag{2}$$

The dropout rates defined as T_a and t_a can be written with this model as:

$$T_a = \sum_{i=1}^{N_y-1} T_a^{(i)} = (T_a + t_a)\sum_{i=0}^{N_y-2} a_i \quad ; \quad t_a = \sum_{n=N_y}^{\infty} T_a^{(i)} = (T_a + t_a)\sum_{i=N_y-1}^{\infty} a_i \tag{3}$$

2.3.2 Mathematical model for the graduation rate

The graduation rate (Van Den Berg, 2005) in the Spanish regulation (Spanish regulation RD 1393, 2007) is the percentage of students relative to their cohort of new income that complete their degree in N_y or N_y+1 years. This is also a statistical input parameter of the model. A more detailed description of the graduation of students, year by year, is again useful for the purposes of this work. The percentage of students graduated after j years will be denoted by $T_g^{(j)}$, with $j \geq N_y$. According the definition, the official graduation rate is the sum of the first two terms of the series. The percentage of students graduated in the year N_y is denoted here by $(1-y)T_g$ (with $0 \leq y \leq 1$) while those graduated in the year N_y+1 are given by the percentage yT_g. A decreasing behavior, statistically characterized, is now assumed starting on the second term. Besides, t_g will represent the *late graduation rate*, which is the percentage of students graduating after N_y+2 or more years. The decreasing series starts at the second term and its sum equals yT_g+t_g. The following formulation is then adopted:

$$T_g^{(N_y)} = (1-y)T_g \quad ; \qquad T_g^{(N_y+1+j)} = \left[yT_g + t_g\right]g_j \quad (j \geq 0) \qquad (4)$$

where g_j are the point probabilities corresponding to each year for the statistical distribution of graduation rates. Denoting by γ to the mean value of the distribution, the following relations must be satisfied:

$$\sum_{i=0}^{\infty} g_j = 1 \quad ; \qquad \sum_{i=0}^{\infty} j \cdot g_j = \gamma \qquad (5)$$

2.3.3 Example of graduation and dropout rates with geometrical decreasing distributions

The discrete probability distribution of geometric type is suitable for statistical modeling of dropout rates and graduation. Firstly, decreasing probabilities result when increasing the current year. This is consistent with the fact that, after a certain point in time, both leaving and graduated students within the same promotion will be reduced year after year. Second, it is a distribution characterized by only one single parameter. This fact contributes to simplify the mathematical formulation. In appendix section A.1 the details of the mathematical formulation used can be found, while throughout the text the model of academic indicators will be illustrated with graphic examples. Fig. 1 shows the model of graduation and dropout of a four-year degree with T_a = 0.2 and T_g = 0.7. Table I summarizes the annual rate of dropout and graduation.

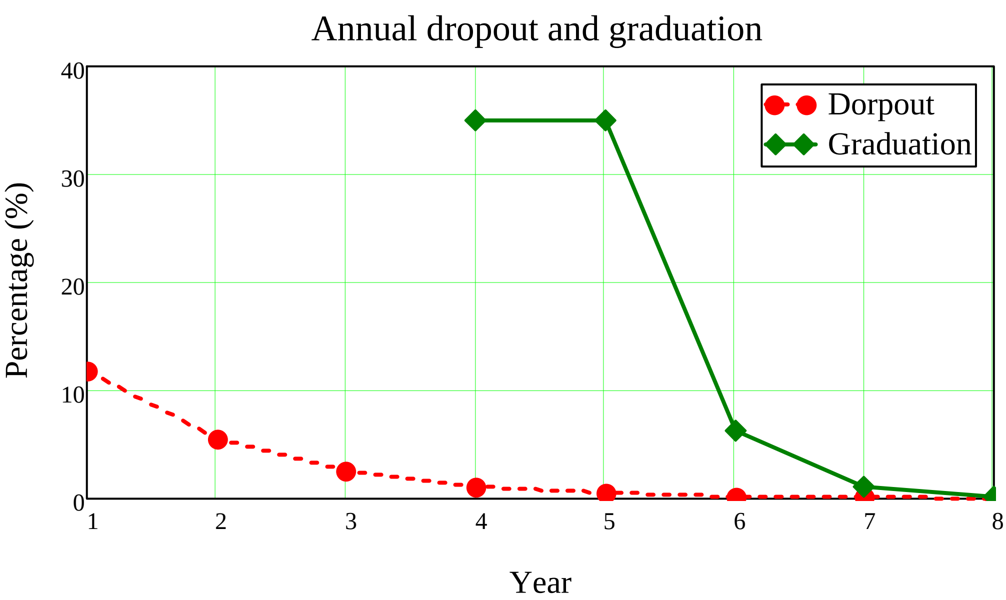

Figure 1. Dropout and graduation model for a four years degree T_a=0,2, T_g=0.7, t_a=0.022, t_g=0 and y=0.5.

In the example $T_g^{(4)}=T_g^{(5)}$, due to the fact of being y=0.5, where $T_g^{(5)}$ is the initial term of the decreasing sequence. This model has been adopted for greater generality, since any other ratio between these first two terms may be considered.

Year:	1	2	3	4	5	6	7	8
Graduation:	-	-	-	35%	35%	6.4%	1.2%	0.21%
Dropout:	11.9%	5.5%	2.6%	1.2%	0.55%	0.26%	0.12%	0.055%

Table 1. Numeric values for the example of Fig. 1

2.3.4 Simple parametric model.

In general, it is desirable to have small rates for the late graduation (t_g) or late dropout (t_a) because they represent the excessive and unnecessary permanence of students in the program. The different components of graduation and dropout, whose sum must be equal to unity, are illustrated in Fig. 2:

$$\left(T_g + t_g\right) + \left(T_a + t_a\right) = 1 \quad (6)$$

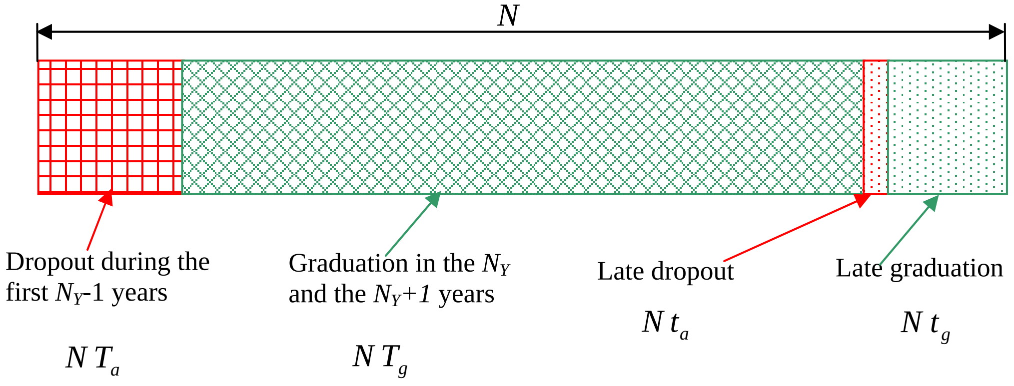

Figure 2. Components of the graduation and dropout rate

Due to equation (6), only three of the four terms can be considered independent variables, because the fourth one is determined by such relation. For convenience, the following three independent variables are proposed for the model:

$$\tau_g = \frac{t_g}{T_g}\,; \quad \tau_a = \frac{t_a}{T_a}\,; \quad k = \frac{\tau_g}{\tau_a} \quad (7)$$

Under normal conditions, both τ_g and τ_a should be small (compared to the unit), which means that most of the students achieve graduation or leave the studies during the first years. This will produce some simplification in the proposed model.

2.4. Duration of studies for graduated students (D_g) and for those leaving the program (D_a)

The mean duration of studies for graduates can be determined as the weighted average of the number of years used by students to achieve graduation. Taking into account the annual graduation rate represented by (4), the duration of the studies is:

$$D_g = \frac{\sum_{j \geq N_y} T_g^{(j)}\, j}{\sum_{j \geq N_y} T_g^{(j)}} \quad (8)$$

Fig. 3 presents the value of (D_g-N_y), which means the excess of duration of studies respect to N_y as a function of the parameter τ_g for different values of y under the statistical geometric model previously described. The details of the mathematical formulation can be found in the appendix section A2. It can be observed that the mean duration is approximately half year more than N_y for moderately small values of τ_g.

For the students that finally leave the program, the mean number of years of permanence can be determined as the weighted average of the number of years in which the students are enrolled before leaving. Taking into account the annual dropout rate represented by (1), the duration of the studies is:

$$D_a = \frac{\sum_{i\geq 1} T_a^{(i)}\, i}{\sum_{i\geq 1} T_a^{(i)}} \tag{9}$$

Fig. 4 illustrates the behavior of D_a as a function of the parameter τ_a for degrees of 3, 4, 5 and 6 years. The details of the formulation are included in appendix section A2. An increasing trend can be observed (as expected) with the number of years and with τ_a, together with an almost linear behavior respect to this last parameter.

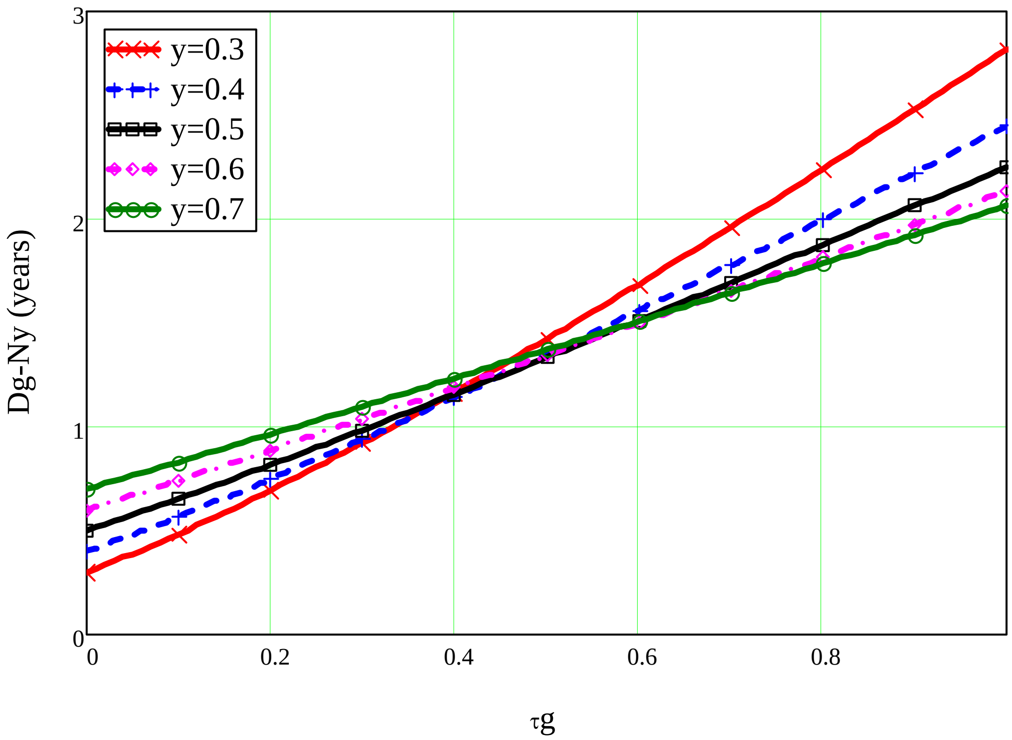

Figure 3. Duration of studies for graduates (excess over N_y)

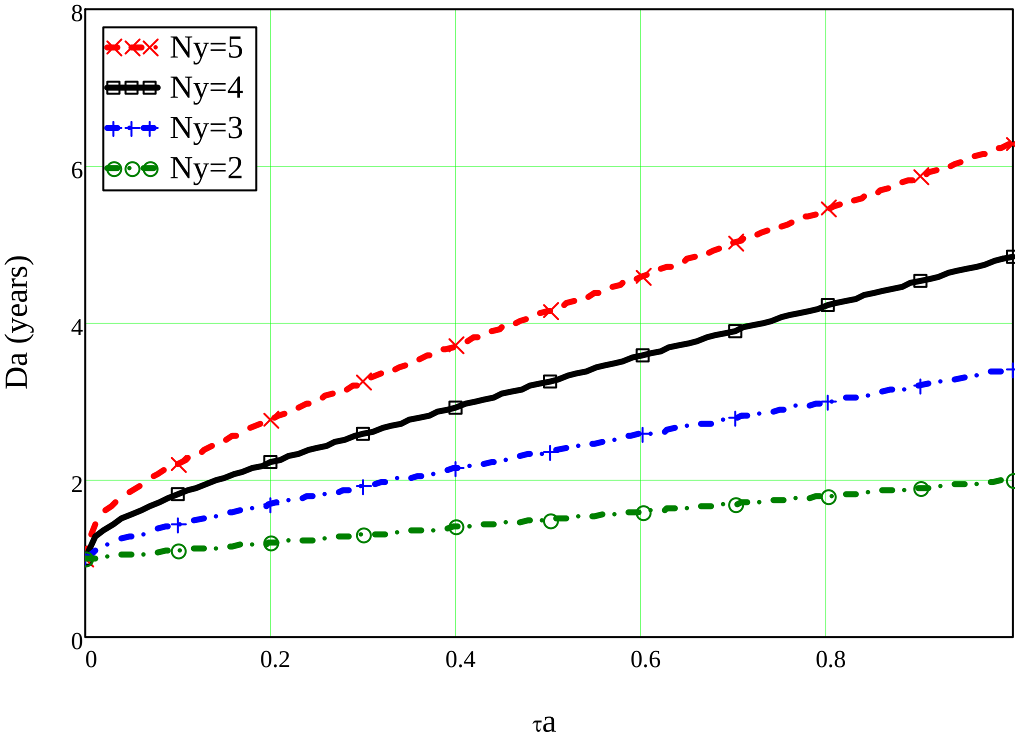

Figure 4. Years of permanence of students that leaves the program

2.5. Number of students enrolled in the program (N_s)

The number of students present in the studies is the key factor for measuring the workload of the faculty staff. If the graduation and dropout rates are considered stationary as well as the number of students of new income, the number of students that are enrolled in the program of studies at any moment will be also stationary. It can be shown that under these conditions the global number of students has two components:

- Those students belonging to the group that graduate. This group represents a percentage of T_g+t_g of the N incoming students and their average duration in the program is D_g.

- Those students belonging to the group that leave the program without graduating after an average of D_a years. This group represents a percentage of T_a+t_a of the incoming N students.

The number of students can be calculated as the sum of both components, as represented in Fig. 5 by the shadowed areas.

$$N_s = N(T_g + t_g)D_g + N(T_a + t_a)D_a \qquad (10)$$

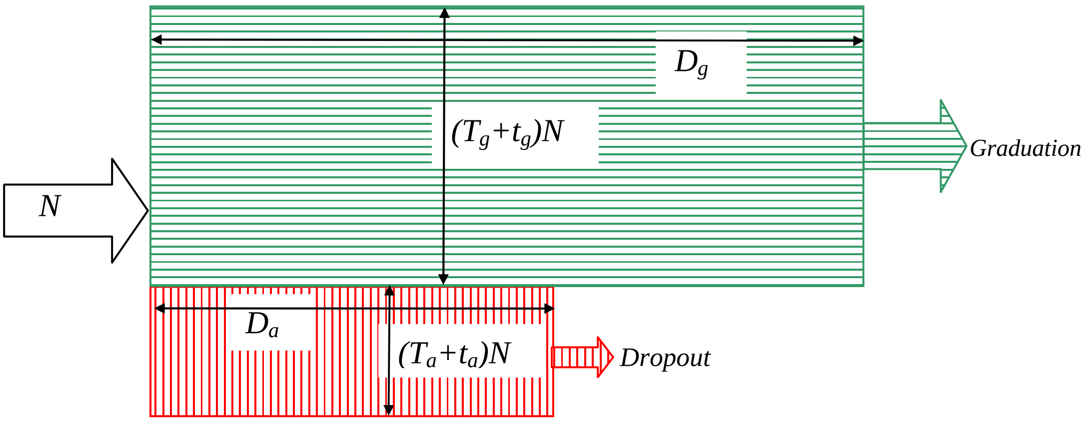

Figure 5. Scheme of the flux of students enrolled in the program of studies and their mean permanence in the studies

Figure 6 shows the dependence of the number of students enrolled (normalized by N) with the dropout rate for different values of the graduation rate when the rest of the independent variables have the reference values ($N_y=4$, $y=0.5$, $k=1$). It can be observed that for the ideal case $T_a=0$ and $T_g=100\%$ the number of students enrolled is 4.5 times N, in other words, $N_s=(N_y+0.5)N$.

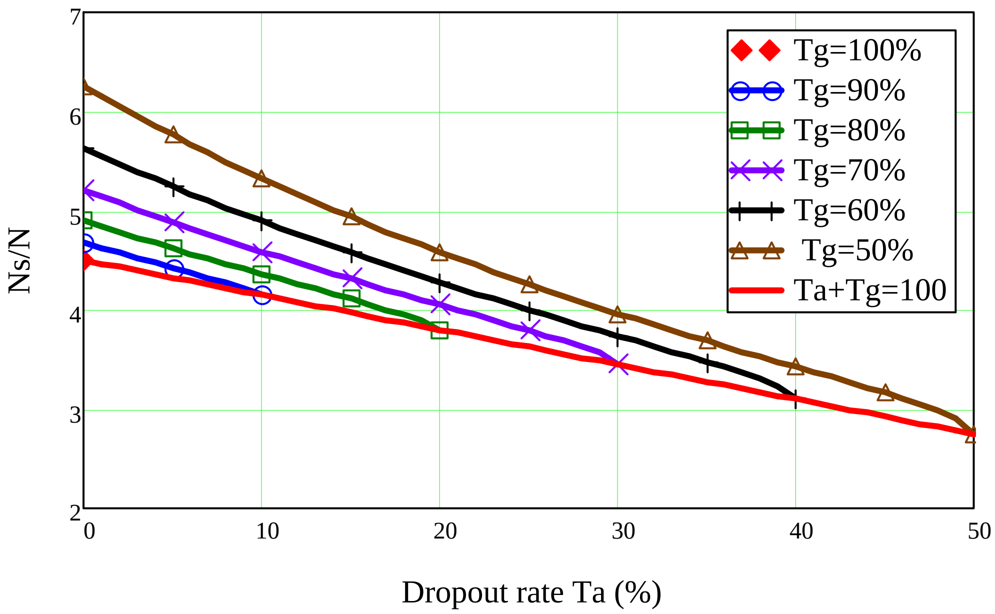

Figure 6. Number of students enrolled for each new incoming student for $N_y=4$, $y=0.5$, and $k=1$ for different values of T_g as a function of T_a.

2.6. Efficiency rate (T_e) and number of credits inscribed (C)

The efficiency rate is defined by Spanish regulations (Spanish regulation RD 1393, 2007) as the ratio between the number of credits in which a group of graduate students have succeeded and the number of credits enrolled by the same group of students along the years required by them for achieving graduation. The group of students is referred to those belonging to the same cohort of new income and finally graduated. Once the number of enrolled students has been determined, the efficiency rate is useful to determine the total number of credits inscribed by all the students in the program. The hypothesis consists of assuming that an average student achieves graduation after inscribing N_c/T_e credits in D_g years in order to complete the degree with the N_c credits necessary for graduation. So, each year the average student of the group that finally achieves graduation (upper part in Figure 5) will be enrolled in $N_c/(T_eD_g)$ credits. An assumption must be made for the students of the group that will leave the studies without achieving graduation. In the absence of specific records, it can be assumed that the total amount of credits enrolled each year is the same for the average students of both groups. Under that assumption, the number of credits inscribed during an academic year by all the students enrolled in the program is:

$$C = N_s \frac{N_c}{T_e D_g} = N \frac{N_c}{T_e}\left[(T_g + t_g) + (T_a + t_a)\frac{D_a}{D_g}\right] \qquad (11)$$

It is interesting to normalize the previous result in order to define the number of credits enrolled related to the number of students of new income and the number of credits of the degree for the case of efficiency rate equal to unity. The following ratio is referred as the normalized number of credits C_N:

$$C_N = \frac{C\,T_e}{N\,N_c} = \left[T_g(1+\tau_g) + T_a(1+\tau_a)\frac{D_a}{D_g}\right] \qquad (12)$$

Fig. 7 shows the C_N ratio for the same example of Fig. 6. This figure shows a strong dependency with the dropout rate with an almost linear behavior.

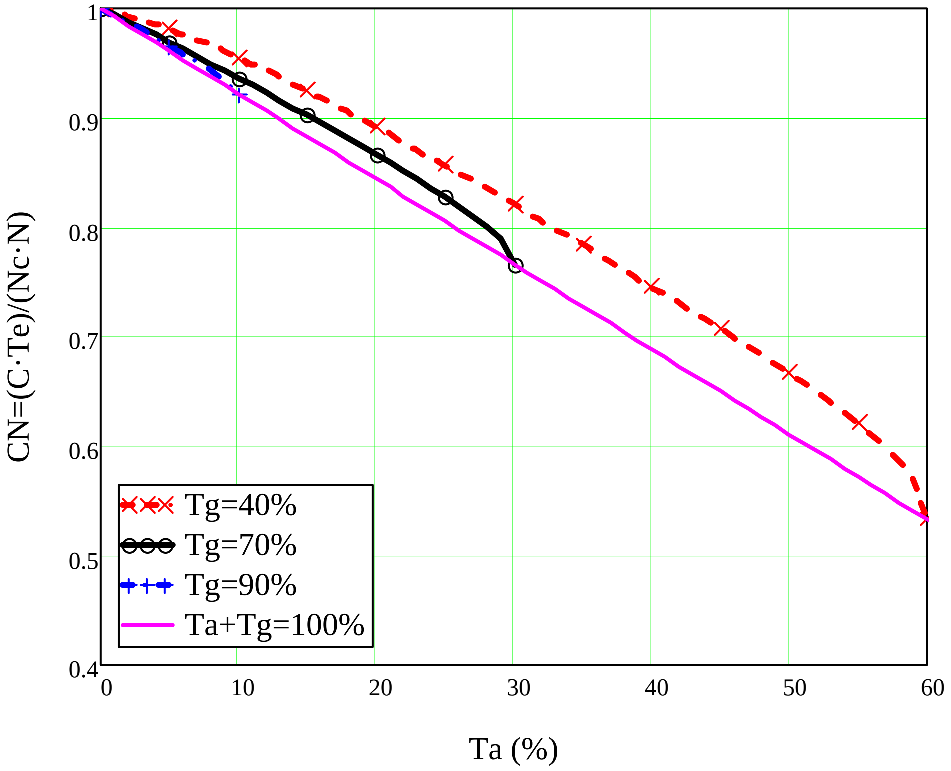

Figure 7. Normalized number of credits inscribed for N_y=4, y=0.5, and k=1 for different values of T_g as a function of T_a.

In the previous examples analyzed, the parameters k and y have been established as constant values. In order to confirm that the dependency with the dropout rate is predominant, an exhaustive parametric study has been developed by taking up to 10,000 randomly generated cases. For each of these cases, random values have been assigned to the independent variables characterizing the mathematical model. Uniform distributions have been considered for the individual generation of random values with the limits shown in Table 2. Besides, a reference value is presented for each parameter for comparison purposes.

Variable	Minimum value	Reference value	Maximum value
N_y	3	4	5
y	0.3	0.5	0.7
k	0.25	1	4
T_a	0%	30%	60%
T_g	40%	70%	100%
T_a+T_g	0%	---	30%

Table 2. Limits and reference values for the parametric study.

Fig. 8 presents the results of the normalized number of credits inscribed in the parametric study. Each dot in the figure represents one of the 10,000 cases considered. As it can be seen in Fig. 8, for zero dropout rate the normalized number of credits enrolled tends to one. An almost linear decreasing behavior depending on T_a can be observed without a very strong dependence on the rest of the parameters. A linear regression of the form (1-0.68·T_a) has been found for best fitting the collection of data points. This linear regression has been also plotted for comparison, showing a high degree of correlation between the randomly selected cases with this linear estimation. As a consequence, the linear estimation is a reasonable criterion for the prediction of the credits inscribed.

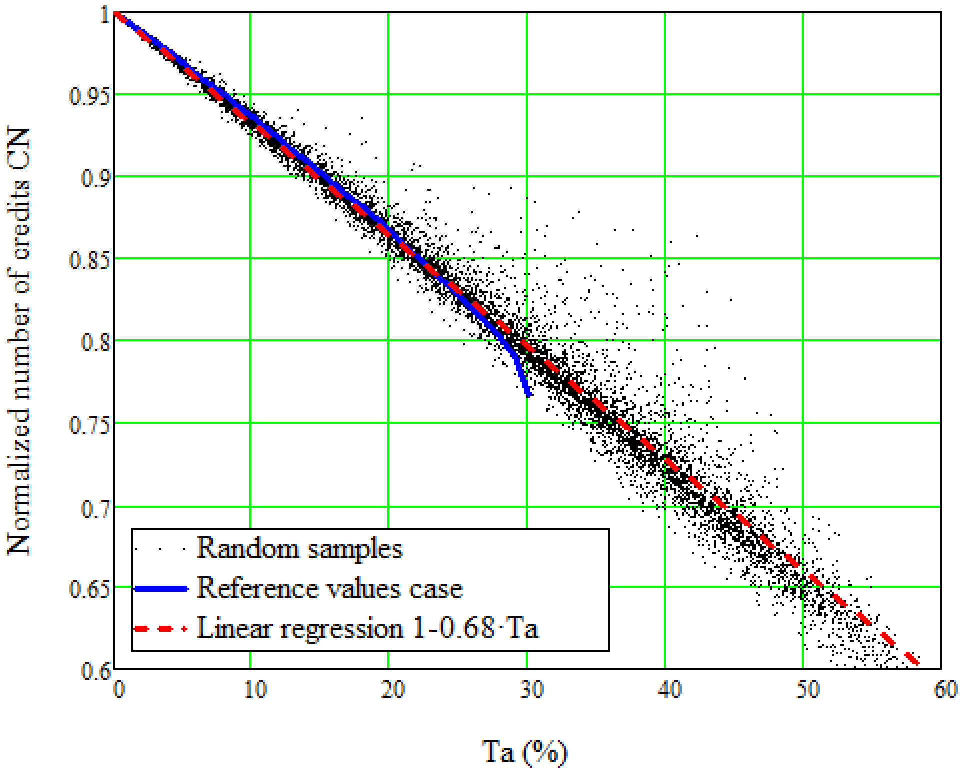

Figure 8. Normalized number of credits inscribed for a number of 10000 cases randomnly selected.

3. PARAMETERS DERIVED FROM REGULATIONS

The internal regulations of universities can define some parameters involving the organization and planning of teaching which are devoted to ensure the quality of the educational process at the same time that the organization scheme is economically feasible. Two parameters will be included here in this group: the number of contact hours per ECTS for students and the average size of groups. On the other hand, labor law establishes the number of teaching hours per year that can be assigned to the faculty staff for regular tuition in classroom, labs, etc.

3.1. Contact hours per ECTS (*H*)

Each ECTS involves between 25 and 30 hours of work by students. These hours include both hours developed in classroom or laboratory under the direct intervention of teachers, as well as the autonomous activity by the student. As already pointed out, in annual average values students take 60 credits a year, representing between 1,500 and 1,800 hours of academic activity. The total amount of contact hours may vary depending on the program or the year/level, but it can be established in margins between 25% and 40% of the total hours. This parameter must be established by the universities.

3.2. Faculty hours (*F*) and average size of groups (*ASG*)

The average size of groups is traditionally defined as the ratio between the enrolled credits by a group of students and the corresponding credits taught by the faculty. Since in the EHEA ECTS credits are reserved exclusively to measure the work load of students, the definition must be adapted to the new context. Instead of talking of credits we will compute hours. As an example, consider a course or class with 45 students enrolled that must attend 60 contact hours at classroom or labs. The activity is distributed into 30 hours in a classroom where the 45 students attend simultaneously and 30 more lab hours where only 15 students attend at the same time. So the students are divided into 3 lab groups. The number of "enrolled hours" by students is 45x60=2700, while the number of hours taught by faculty is 30x1+30x3=120. The average size of groups for that course is 2700/120=22.5, which represents the average group considering classroom and lab hours.

The contact hours received by students is the number of ECTS enrolled times *H*, which can be compared to the number of lecture hours by the faculty. The *ASG* is then defined as follows:

$$ASG = \frac{C\,H}{F} \tag{13}$$

where *F* is the total number of hours taught by the faculty. Although there are other kind of working hours (tutoring, lesson preparation, etc.), the number of hours considered here is only the co-called 'contact' hours because they determine staffing needs. If the *ASG* is established by regulations, the necessary faculty hours is determined by simply inverting (13):

$$F = \frac{C\,H}{ASG} = N\,\frac{N_c}{T_e}\,\frac{H}{ASG}\left[T_g\,(1+\tau_g) + T_a\,(1+\tau_a)\,\frac{D_a}{D_g}\right] \tag{14}$$

Actually, the ratio *ASG/H* can be regulated by the university in order to establish the proportionality between credits inscribed and staffing needs in contact hours.

If the linear approximation is assumed, the faculty hours are:

$$F = N\,\frac{N_c}{T_e}\,\frac{H}{ASG}\left[1 - 0.68\,T_a\right] \tag{15}$$

The above expression is extremely simple and useful for the human-resource planning of faculties in universities. The number of required teaching hours can be set depending on the recorded values or extracted statistics (N, N_c , *Ta*) and, secondly, on the values that the university can set as targets to be met (T_e , *H/ASG*).

3.3. Annual teaching capacity of the faculty (C_f), number of faculty staff (N_f) and ratio students/teachers (R)

The number of hours C_f that educators must teach in activities as those reserved for contact hours is derived from regulations. In this model, the reference is the full-time activity for both students and faculty. The dimension of the faculty staff derived from the model must be considered a full-time equivalent. The total number of faculty individuals is directly derived from the hours of contact activity and the annual capacity:

$$N_f = \frac{P}{C_f} \approx \frac{N}{C_f}\,\frac{N_c}{T_e}\,\frac{H}{ASG}\left[1 - 0.68\,T_a\right] \tag{16}$$

The ratio between the number of students enrolled and the number of faculty individuals is:

$$R = \frac{N_s}{N_f} = \frac{C_f}{N_c}\left(T_e\ D_g\right)\frac{ASG}{H} \qquad (17)$$

The ratio *RSF* has been written as the product of three factors. The first one (C_f/N_c) usually depends on national regulations. As an example, according to Spanish regulations, the "Grado" degree consists of N_c=240 ECTS while the number of hours assigned to full-time faculty is 240 per year. The second factor (T_eD_g) is related to academic features of the degree, but under normal circumstances it is close to the number of years of the degree. The last one (*ASG/H*) contains two key parameters that the university can control to ensure the quality and financial feasibility of the higher education

Fig. 9 presents, as an example, the ratio *R* in terms of the average size of groups (*ASG*) and the contact hours per ECTS (*H*) for a four year degree, assuming k=1, y=0.5, T_a+T_g=90%, T_e=90%, and (C_f/N_c)=1.

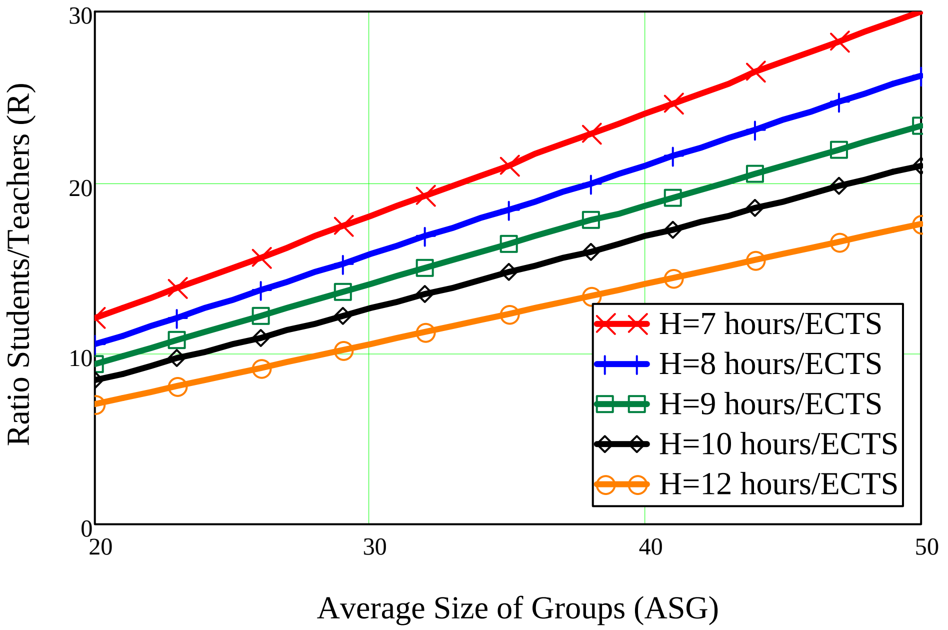

Figure 9. Students/Teachers ratio (R) Ratio for a four year degree with k=1, y=0.5, T_a+T_g=90%, T_e=90%, and (C_f/N_c)=1

4. CONCLUSIONS

A mathematical model has been presented which allows the prediction of the dimension of faculty staff necessary for the development of a study program or degree in higher education. The model starts with a small number of input parameters taken from statistical records, according to the official definition of the indicators used in the EHEA and some parameters derived from university, regional or national regulations. Using some simplifications which are based on experience, an explicit formulation has been developed which can be useful for the bodies of government of the universities. The formulae include some parameters which depend on the organization and planning of teaching, so the process can be controlled to ensure a compromise between the criteria related to the quality of education and the feasibility of the solution from the financial point of view. The exhaustive parametric study made by randomly varying some parameters of the model without significant variation of the final result allows the corroboration of the robustness of the model.

ACKNOWLEDGMENT

This work has been supported by the University of Vigo through the Office of the Commissioner for New Projects.

REFERENCES

Bromwich, T. J. I'A. and MacRobert, T. M. (1991). *An Introduction to the Theory of Infinite Series, 3rd ed*. New York: AMS Chelsea Publishing.

Cabrera, L., Tomás, J., Álvarez, P. and Gonzalez, M. (2006). "El problema del abandono de los estudios universitarios". RELIEVE, v. 12, n. 2, p. 171-203. http://www.uv.es/RELIEVE/v12n2/RELIEVEv12n2_1.htm

European Commision. (2004). *European Credit Transfer and Accumulation System (ECTS) - Key Features*. Luxembourg: Office for Official Publications of the European Communities.

García Pino, A. (2012). "Measuring the workload of teachers in the European Higher Education". *Proceedings of INTED 2012*. Valencia 5-7 Mar. 2012.

Spanish Regulation, RD 1393. (2007). RD 1393/2007 de 29 de octubre por el que se establece la ordenación de las enseñanzas universitarias oficiales. BOE 260, 18770, 30/10/2007.

Tovar, E.; Plaza, I., Castro, M.; Llamas, M., Arcega, F., Jurado, F., Mur, F., Sanchez, J. A., Falcone, F., and Dominguez, M. (2007). "Modeling the best practices towards the adaptation to the European credit transfer system in technical degrees within the IEEE ES chapter". *Proceedings of Frontiers in Education Conference*, 2007, pp. 538-544.

Tinto, Vincent. (1975). "Dropout from Higher Education: A Theoretical Synthesis of Recent Research". *Review of Educational Research*, Winter 1975 45: 89-125, doi:10.3102/00346543045001089.

Van Den Berg, MN and Hofman WHA. (2005). "Student success in university education: A multi-measurement study of the impact of student and faculty factors on study progress". *Higher Education*, vol. 50, no. 3, pp.413-446. doi: 10.1007/s10734-004-6361-1

Weisstein, Eric W. Geometric Distribution. From MathWorld--A Wolfram Web Resource. http://mathworld.wolfram.com/GeometricDistribution.html

APPENDIX. MATHEMATICAL FORMULATION

A.1 Geometric probability distributions for graduation and dropout rates.

The discrete probability distribution of the geometric type is based on only one single parameter and involves decreasing probability for increasing the number of years. The probability density of the geometric type (Bromwich, 2005) (Weisstein) is determined by the probability that the event occurs at time i :

$$P(i) = p\,(1-p)^{i} \quad i \geq 0 \tag{18}$$

here p is the characteristic parameter of the distribution. The mean value μ is directly related to the parameter p:

$$\mu = \frac{1-p}{p} = \frac{1}{p} - 1 \tag{19}$$

The proposed distribution determines the probability that the event occurs at time i or before:

$$P(\leq i) = \sum_{j=0}^{i} p\,(1-p)^{j} = 1-(1-p)^{i+1} \tag{20}$$

The above equation is derived from the properties of the geometric progression series (Bromwich, 2005).

The parameter of the statistical distribution a_i used for dropout modeling can be obtained from (3) by taking into account (20):

$$T_a = (T_a + t_a)\sum_{i=0}^{N_y-2} a_i (T_a + t_a) = \sum_{i=0}^{N_y-2} p_a (1-p_a)^{i} = (T_a + t_a)\left[1-(1-p_a)^{N_y-1}\right] \tag{21}$$

Then, the characteristic parameter p_a of the geometric dropout distribution can be obtained as:

$$p_a = 1 - \sqrt[(N_y-1)]{\frac{t_a}{T_a + t_a}} = 1 - \sqrt[(N_y-1)]{\frac{\tau_a}{1+\tau_a}} \tag{22}$$

The mean value is given by (19):

$$\alpha = \mu_a = \frac{1}{p_a} - 1 = \frac{1}{1 - \sqrt[(N_y-1)]{\frac{\tau_a}{1+\tau_a}}} - 1 \tag{23}$$

For the graduation rate the corresponding parameter p_g, for characterizing the statistical distribution g_j, when it fits the geometric type can be deducted from (4) by considering the graduation at year (N_y+1):

$$T_g^{N_y+1} = \left(yT_g + t_g\right) g_0 \Rightarrow yT_g = \left[yT_g + t_g\right] p_g \tag{24}$$

From above equation the p_g can be solved, and then the mean value of the distribution γ can be obtained too:

$$p_g = \frac{yT_g}{yT_g + t_g} = \frac{y}{y + \tau_g} \tag{25}$$

$$\gamma = \frac{1}{p_g} - 1 = \frac{\tau_g}{y} = \frac{t_g}{yT_g} \qquad (26)$$

A.2 Duration of studies

The years of permanence in studies for those students leaving the degree without success can be deducted by introducing the annual dropout rates (1) into (9):

$$D_a = \frac{\sum_{i\geq 1} i \cdot T_a^{(i)}}{\sum_{i\geq 1} T_a^{(i)}} = \frac{\sum_{i\geq 0} (i+1) \cdot T_a^{(i+1)}}{\sum_{i\geq 0} T_a^{(i+1)}} = \frac{(T_a + t_a)\sum_{i\geq 0}(i+1)\cdot a_i}{(T_a + t_a)\sum_{i\geq 0} a_i} \qquad (27)$$

After removing the common factor *(T_a+t_a)* and by making use of the properties of the statistical distribution given by (2), it follows:

$$D_a = \frac{\sum_{i\geq 0}(i+1)\cdot a_i}{\sum_{i\geq 0} a_i} = \frac{\sum_{i\geq 0} i\cdot a_i + \sum_{i\geq 0} a_i}{\sum_{i\geq 0} a_i} = \frac{\alpha+1}{1} = \alpha + 1 \qquad (28)$$

The duration of the graduate studies may be obtained similarly, although the deduction is more laborious. The introduction of the graduation rates given by equation (4) in the definition (8) allows writing:

$$D_g = \frac{\sum_{j\geq N_y} j\cdot T_g^{(j)}}{\sum_{j\geq N_y} T_g^{(j)}} = \frac{N_y(1-y)T_g + (yT_g + t_g)\sum_{j\geq 0}(N_y + 1 + j)\cdot g_i}{(1-y)T_g + (yT_g + t_g)\sum_{i\geq 0} g_i} \qquad (29)$$

The numerator can be decomposed as follows:

$$D_g = \frac{N_y(1-y)T_g + (yT_g + t_g)\left[(N_y+1)\sum_{j\geq 0} g_i + \sum_{j\geq 0} j\cdot g_i\right]}{(1-y)T_g + (yT_g + t_g)\sum_{i\geq 0} g_i} \qquad (30)$$

Using now the properties of the statistical distribution g_i given by equation (5), it follows:

$$D_g = \frac{N_y(1-y)T_g + (yT_g + t_g)(N_y + 1 + \gamma)}{(1-y)T_g + (yT_g + t_g)} = \frac{N_y(T_g + t_g) + (yT_g + t_g)(1+\gamma)}{T_g + t_g} \qquad (31)$$

The above expression can be written more simply in terms of the coefficient τ_g solving the excess of studies duration over N_y:

$$D_g - N_y = \frac{(y+\tau_g)(1+\gamma)}{1+\tau_g} \qquad (32)$$

Particularizing the results of equations (28) and (32) for geometric distributions by entering values α and γ from expressions (23) and (26), respectively, it follows:

$$D_a = \alpha + 1 = \frac{1}{p_a} = \frac{1}{1 - \sqrt[(N_y-1)]{\frac{\tau_a}{1+\tau_a}}} \qquad (33)$$

$$D_g - N_y = \frac{(y+\tau_g)(1+\gamma)}{1+\tau_g} = \frac{(y+\tau_g)^2}{y \cdot (1+\tau_g)} \qquad (34)$$

Figures 4 and 3 provide a graphic representation of these last two expressions.