

Document downloaded from:

<http://hdl.handle.net/10251/36652>

This paper must be cited as:

Cortés López, JC.; Sánchez Sánchez, A.; Santonja, FJ.; Villanueva Micó, RJ. (2013). Non-parametric probabilistic forecasting of academic performance in Spanish high school using an epidemiological modelling approach. *Applied Mathematics and Computation*. 221(15):648-661. doi:10.1016/j.amc.2013.06.070.



The final publication is available at

<http://dx.doi.org/10.1016/j.amc.2013.06.070>

Copyright Elsevier

# Non-parametric probabilistic forecasting of academic performance in Spanish high school using an epidemiological modelling approach

Juan-C. Cortés<sup>1</sup>, Almudena Sánchez-Sánchez<sup>1</sup>, Francisco-J. Santonja<sup>1</sup>,  
Rafael-J. Villanueva<sup>1</sup>

<sup>a</sup>*Instituto Universitario de Matemática Multidisciplinar, Building 8G, 2<sup>nd</sup> floor,  
Universitat Politècnica de València, 46022 Valencia (Spain)*

<sup>b</sup>*Departamento de Estadística e Investigación Operativa, Universidad de Valencia, 46100  
Burjassot, Valencia (Spain)*

---

## Abstract

Academic underachievement is a concern of paramount importance in Europe, and particularly in Spain, where around of 30% of the students in the last two courses in high school do not achieve the minimum knowledge academic requirement. In order to analyse this problem, we propose a mathematical model via a system of ordinary differential equations to study the dynamics of the academic performance in Spain. Our approach is based on the idea that both, good and bad study habits, are a mixture of personal decisions and influence of classmates. Moreover, in order to consider the uncertainty in the estimation of model parameters, a bootstrapping approach is employed. This technique permits to forecast model trends in the next few years using confidence intervals. Unfortunately, the obtained results do not suggest improvement in academic performance for the coming years.

*Keywords:* Academic underachievement, Modelling, Transmission dynamics, Uncertainty, Bootstrapping, Prediction.

---

## 1. Introduction

During the last years Spanish authorities have carried out several changes in the educational laws looking for an improvement in academic results [1, 2].

---

\*Corresponding Author: Phone: 00-34-963-877000 (ext. 88289).

These legislative measures have been focused on the educational levels of Compulsory Secondary Education (in the Spanish terminology, Educación Secundaria Obligatoria (ESO)) and the last two courses of high school (in the Spanish terminology *Bachillerato*) that correspond to students between 12–16 and 16–18 years old, respectively. Despite of these efforts, the rates of academic underachievement in these educational stages are still at very worrying levels, about 30% of the pupils [3]. Although less pronounced, these figures are also alarming in many other countries of the European Union [4, 5, 6, 7].

The concern about the high level of academic underachievement in Spain is completely justified, not only by the high rates but also by the negative effects on the country's economic development [8], especially in the unemployment and its serious consequences. This issue is of primer importance in the current context of economic crisis affecting particularly Spain.

Nowadays, the job opportunities of people depend on their qualification, their ability to acquire, use and interpret the information, including their skills to adapt the new knowledge to a very demanding and competitive society in constant change. In order to acquire them, students go to schools first and high schools later, learning the contents determined in the current legislation. In Spain, the end of the high school courses corresponds to *Bachillerato* and it represents a key-point as we will see later [9].

The access to this academic educational level takes place after finishing the Compulsory Secondary Education (ESO). According to the legislation currently into force [1], the main objective of the academic level of *Bachillerato* is to provide 16–18 years old Spanish students a proper educational training to consolidate the intellectual maturity of the pupils, increasing their specific knowledge as well as boosting the development of abilities that help them to join up either the labor market or higher studies. For all these reasons, this educational level is considered a milestone to students because it represents a period to make important decisions about academic and professional future.

In Spain, students can especialise in three different *Bachillerato* branches: Arts; Sciences and Technology; Humanities and Social Sciences (see Figure 1). In September, a student of First Stage of *Bachillerato* will transit into the Second Stage of *Bachillerato* if she/he has passed successfully all the subjects except, maybe, at most two out of a total of ten. In order to get the *Bachillerato* degree, it is necessary to pass all the subjects of both

*Bachillerato* Stages [2].

### **Caption Figure 1**

Structure of the Spanish educational system for students aged 12-18. Most of the students follow the academic path ESO + *Bachillerato*.

According to the different learning theories [1, 1], in particular, the Vygotskian perspective [1, 1, 1] and the recent studies published [1], habits and behavior may be socially transmitted, in particular, academic and study habits. Taking into account this approach, in this paper we propose to model the evolution of the academic performance in the educational level of *Bachillerato* in Spain using modelling techniques in mathematical epidemiology. Some examples of analogous situations using type-epidemiological mathematical models are encountered in public health, obesity [1, 1], alcoholism [1], drug abuse [2], shopaholism [2], spread of ideas [2], evaluation of law effects on societies [2], and so on.

A first paper following this approach has been published in [2]. In this contribution, bad academic habits are considered as behavior susceptible to be transmitted among students, mainly between students in the same academic level, on a gender-structured model.

In this paper, we propose a gender-and-course-structured model where we consider, apart from the spread of bad academic habits, the spread of good ones (positive transmission) from students belonging to the promotable group to the non-promotable one. Notice that this additional issue is based on pedagogical strategies that consider mixing groups of students with bad and good academic results in order to induce improvement of them (Educational Inclusion) [2]. Furthermore, we include the estimation of the abandon rates. Abandon is an important aspect still under debate in the pedagogical area which quantification is difficult. We have made a decision in order to include this issue in the model and this is to consider *abandon* when, during the academic year, the student leaves the academic system.

Once the model is stated, we will be able to monitor the promoted and graduated students. Other new contribution is the introduction of uncertainty in the obtention of the value of the parameters of our model which will allow us to predict the evolution of the academic performance in specific confidence intervals.

The proposed approach will allow us to understand better the mechanism behind the academic performance as well as to predict how things will evolve in the Spanish *Bachillerato* over the next few years and this way, to provide relevant information to make appropriate decisions to policymakers.

This paper is organised as follows. In Section 2, we build and introduce the model. Once the model is stated, unknown parameters are estimated. We present the parameter estimation procedure in Section 3. In Section 4, uncertainty in model parameters is introduced using bootstrapping technique and predictions over the next few years about academic performance in *Bachillerato* are presented. Section 5 deals with the quantification and analysis of abandon. Conclusions are given in Section 6.

## 2. The epidemiological-mathematical model

### 2.1. Available data

The available data that we have considered in this paper correspond to the academic results belonging to the students of the First and Second Stage of *Bachillerato* during the academic years from 1999 – 2000 to 2008 – 2009, in both, state and private high schools all over Spain. For each academic year, Table 1 collects [2]:

#### Caption Table 1

The available data corresponding to the First and Second Stage of *Bachillerato*, in both, state and private high schools all over Spain from academic year 1999 – 2000 to 2008 – 2009. Each row shows the percentage of Girls and Boys who promote ( $G_i | B_i$ ) and do not promote ( $\overline{G}_i | \overline{B}_i$ ) for each level  $i = 1, 2$  over the total Spanish *Bachillerato* students.

- The percentage of girls|boys who promote ( $G_i|B_i$ ), that is, girls and boys with study behavior good enough to promote to the Second Stage, if they are in the First Stage ( $i = 1$ ), or to obtain the *Bachillerato* degree, if they are in the Second Stage ( $i = 2$ ), at the end of each course.
- The percentage of girls|boys who do not promote ( $\overline{G}_i|\overline{B}_i$ ), i.e., girls and boys with study habits that do not allow them to reach the minimum

knowledge academic requirement to promote to Second Stage, if they are in the First Stage ( $i = 1$ ), or to obtain their *Bachillerato* degree if they are in the Second Stage ( $i = 2$ ), at the end of each course.

## 2.2. Model building

We build our mathematical model following an epidemiological approach considering as main idea that the academic performance of a student, Girl (G) or Boy (B), is a mixture of her/his own study habits and the study habits, good or bad, of their classmates. In our model, we assume that the transmission of good and bad academic habits is caused by the social contact between students who belong to the same academic level [1, 1, 1, 2].

The subpopulations of the model will be (time  $t$  in years):

- $G_i = G_i(t)$  is the number of girls who are in condition to promote at time instant  $t$ , for  $i = 1, 2$ .
- $B_i = B_i(t)$  is the number of boys who are in condition to promote at time instant  $t$ , for  $i = 1, 2$ .
- $\overline{G}_i = \overline{G}_i(t)$  is the number of girls who are not in condition to promote at time instant  $t$ , for  $i = 1, 2$ .
- $\overline{B}_i = \overline{B}_i(t)$  is the number of boys who are not in condition to promote at time instant  $t$ , for  $i = 1, 2$ .

Furthermore, we consider the following assumptions to build the model:

- Let us assume homogeneous population mixing, i.e., each student can contact with any other student [2].
- *Negative autonomous decision*: For each *Bachillerato* academic level  $i = 1, 2$ , students belonging to the promotable groups  $G_i$  or  $B_i$ , may change their personal habit towards the study and this change may lead the students to obtain bad academic results, moving to  $\overline{G}_i$  or  $\overline{B}_i$ . We assume that this transition is proportional to the number of pupils in  $G_i$  and  $B_i$ , and it is modeled by the linear terms  $\alpha_i^G G_i$  and  $\alpha_i^B B_i$ . According to educational experts, it is assumed that the academic attitude is different in the same educational level depending on gender: girls are usually more responsible for their academic performance than

boys in both the First and Second Stage of *Bachillerato* [2]. This leads us to suppose the following restrictions:

$$\alpha_1^G < \alpha_1^B, \alpha_2^G < \alpha_2^B. \quad (1)$$

In addition we will assume that:

$$\alpha_1^G > \alpha_2^G, \alpha_1^B > \alpha_2^B, \quad (2)$$

because students in the Second Stage are more mature than their mates in the First Stage [2].

- *Negative habits transmission:* For each *Bachillerato* academic level  $i = 1, 2$ , students in  $G_i$  or  $B_i$  may move to the non-promotable group,  $\overline{G}_i$  or  $\overline{B}_i$  respectively, due to the negative influence transmitted in the encounters between students (girls and boys) in the non-promotable group in the same academic level. Hence, these transitions are modeled by the nonlinear terms  $\beta_i^{GG}G_i\overline{G}_i + \beta_i^{GB}G_i\overline{B}_i$  and  $\beta_i^{BG}B_i\overline{G}_i + \beta_i^{BB}B_i\overline{B}_i$ , where  $\beta_i^{GG}$ ,  $\beta_i^{GB}$ ,  $\beta_i^{BG}$  and  $\beta_i^{BB}$  are the corresponding transmission rates where the first letter in the superindexes denotes the group susceptible to acquire the bad study habit and the second one denotes the group that transmit the bad study habit,  $i = 1, 2$ , at time  $t$ . All specific factors and social encounters involved in the transmission of the bad academic habits are embedded in  $\beta$  parameters.
- *Positive autonomous decision:* For each *Bachillerato* academic level  $i = 1, 2$ , students belonging to the non-promotable group  $\overline{G}_i$  or  $\overline{B}_i$ , may change their personal behavior towards their study habits and this change may lead the students to improve their academic results, moving to  $G_i$  or  $B_i$ . We assume that this transition is proportional to the number of pupils in  $\overline{G}_i$  and  $\overline{B}_i$ , and it is modeled by the linear terms  $\gamma_i^G\overline{G}_i$  and  $\gamma_i^B\overline{B}_i$ .
- *Positive habits transmission:* For each *Bachillerato* academic level  $i = 1, 2$ , students in  $\overline{G}_i$  or  $\overline{B}_i$  may move to the promotable group,  $G_i$  or  $B_i$  respectively, due to the positive influence transmitted in the encounters between students (girls and boys) in the promotable group in the same academic level. Hence, these transitions are modeled by the nonlinear terms  $\delta_i^{GG}\overline{G}_iG_i + \delta_i^{GB}\overline{G}_iB_i$  and  $\delta_i^{BG}\overline{B}_iG_i + \delta_i^{BB}\overline{B}_iB_i$ , where  $\delta_i^{GG}$ ,  $\delta_i^{GB}$ ,

$\delta_i^{\overline{BG}}$  and  $\delta_i^{\overline{BB}}$  are the corresponding transmission rates where the first letter in the superindexes denotes the group susceptible to acquire the good study habit and the second one denotes the group that transmit the good study habit,  $i = 1, 2$ , at time  $t$ . All specific factors and social encounters involved in the transmission of the good academic habits are embedded in  $\delta$  parameters.

- *Passing courses and graduation:* According to the Spanish educational law, all the students in  $G_1$  and  $B_1$ , in September, transit automatically to  $G_2$  and  $B_2$ , respectively. Analogously, all the students in  $G_2$  and  $B_2$  will graduate in September. These transitions are modeled by  $\varepsilon G_1, \varepsilon G_2, \varepsilon B_1, \varepsilon B_2$ , where

$$\varepsilon = \begin{cases} 1 & \text{if } \frac{9}{12} + j \leq t \leq \frac{10}{12} + j, \\ 0 & \text{otherwise,} \end{cases}$$

where  $j = 0, 1, \dots, 9$  correspond to academic years 1999–2000, 2000–2001,  $\dots$ , 2008–2009, respectively.

- *Abandon:* As we said previously, this new formulation of the model considers non-completion of *Bachillerato*. For each *Bachillerato* academic level  $i = 1, 2$ , a proportion of the students in  $\overline{G}_i$  or  $\overline{B}_i$  with bad academic results may leave the studies by autonomous decision. This situation is modeled by the linear terms  $\eta_i^G \overline{G}_i$  and  $\eta_i^B \overline{B}_i$ . We also assume that these transitions are proportional to the number of pupils in  $\overline{G}_i$  and  $\overline{B}_i$ .
- *Access:* New students enter into the *Bachillerato* in the month of September in the promotable group, both girls and boys. It is modelled by the functions

$$\sigma^G = \begin{cases} \tau^G & \text{if } \frac{9}{12} + j \leq t \leq \frac{10}{12} + j, \\ 0 & \text{otherwise,} \end{cases}$$

$$\sigma^B = \begin{cases} \tau^B & \text{if } \frac{9}{12} + j \leq t \leq \frac{10}{12} + j, \\ 0 & \text{otherwise,} \end{cases}$$



where  $j = 0, 1, \dots, 9$  correspond to academic years 1999–2000, 2000–2001,  $\dots$ , 2008–2009, respectively, and  $\tau^G$  and  $\tau^B$  to be determined.

Thus, under the above assumptions we build the nonlinear system of ordinary differential equations (3) in order to describe the dynamics of *Bachillerato* students academic performance, where the unknown functions are  $G_i = G_i(t)$ ,  $B_i = B_i(t)$ ,  $\bar{G}_i = \bar{G}_i(t)$  and  $\bar{B}_i = \bar{B}_i(t)$  ( $t$  denotes time in years). For the sake of clarity, each equation is written in three lines, in the first one the linear terms, in the second one the non-linear terms related to the transmission of bad study habits and in the third line the non-linear terms related to the transmission of good study habits.

$$\begin{aligned}
G_1'(t) &= \sigma^G - \varepsilon G_1(t) - \alpha_1^G G_1(t) + \gamma_1^G \bar{G}_1(t) \\
&\quad - \beta_1^{G\bar{G}} G_1(t) \frac{\bar{G}_1(t)}{T(t)} - \beta_1^{G\bar{B}} G_1(t) \frac{\bar{B}_1(t)}{T(t)} \\
&\quad + \delta_1^{\bar{G}G} \bar{G}_1(t) \frac{G_1(t)}{T(t)} + \delta_1^{\bar{G}B} \bar{G}_1(t) \frac{B_1(t)}{T(t)}, \\
\bar{G}_1'(t) &= \alpha_1^G G_1(t) - \gamma_1^G \bar{G}_1(t) - \eta_1^G \bar{G}_1(t) \\
&\quad + \beta_1^{G\bar{G}} G_1(t) \frac{\bar{G}_1(t)}{T(t)} + \beta_1^{G\bar{B}} G_1(t) \frac{\bar{B}_1(t)}{T(t)} \\
&\quad - \delta_1^{\bar{G}G} \bar{G}_1(t) \frac{G_1(t)}{T(t)} - \delta_1^{\bar{G}B} \bar{G}_1(t) \frac{B_1(t)}{T(t)}, \\
G_2'(t) &= \varepsilon G_1(t) - \varepsilon G_2(t) - \alpha_2^G G_2(t) + \gamma_2^G \bar{G}_2(t) \\
&\quad - \beta_2^{G\bar{G}} G_2(t) \frac{\bar{G}_2(t)}{T(t)} - \beta_2^{G\bar{B}} G_2(t) \frac{\bar{B}_2(t)}{T(t)} \\
&\quad + \delta_2^{\bar{G}G} \bar{G}_2(t) \frac{G_2(t)}{T(t)} + \delta_2^{\bar{G}B} \bar{G}_2(t) \frac{B_2(t)}{T(t)}, \\
\bar{G}_2'(t) &= \alpha_2^G G_2(t) - \gamma_2^G \bar{G}_2(t) - \eta_2^G \bar{G}_2(t) \\
&\quad + \beta_2^{G\bar{G}} G_2(t) \frac{\bar{G}_2(t)}{T(t)} + \beta_2^{G\bar{B}} G_2(t) \frac{\bar{B}_2(t)}{T(t)} \\
&\quad - \delta_2^{\bar{G}G} \bar{G}_2(t) \frac{G_2(t)}{T(t)} - \delta_2^{\bar{G}B} \bar{G}_2(t) \frac{B_2(t)}{T(t)}, \\
B_1'(t) &= \sigma^B - \varepsilon B_1(t) - \alpha_1^B B_1(t) + \gamma_1^B \bar{B}_1(t) \\
&\quad - \beta_1^{B\bar{G}} B_1(t) \frac{\bar{G}_1(t)}{T(t)} - \beta_1^{B\bar{B}} B_1(t) \frac{\bar{B}_1(t)}{T(t)} \\
&\quad + \delta_1^{\bar{B}G} \bar{B}_1(t) \frac{G_1(t)}{T(t)} + \delta_1^{\bar{B}B} \bar{B}_1(t) \frac{B_1(t)}{T(t)}, \\
\bar{B}_1'(t) &= \alpha_1^B B_1(t) - \gamma_1^B \bar{B}_1(t) - \eta_1^B \bar{B}_1(t) \\
&\quad + \beta_1^{B\bar{G}} B_1(t) \frac{\bar{G}_1(t)}{T(t)} + \beta_1^{B\bar{B}} B_1(t) \frac{\bar{B}_1(t)}{T(t)} \\
&\quad - \delta_1^{\bar{B}G} \bar{B}_1(t) \frac{G_1(t)}{T(t)} - \delta_1^{\bar{B}B} \bar{B}_1(t) \frac{B_1(t)}{T(t)}, \\
B_2'(t) &= \varepsilon B_1(t) - \varepsilon B_2(t) - \alpha_2^B B_2(t) + \gamma_2^B \bar{B}_2(t) \\
&\quad - \beta_2^{B\bar{G}} B_2(t) \frac{\bar{G}_2(t)}{T(t)} - \beta_2^{B\bar{B}} B_2(t) \frac{\bar{B}_2(t)}{T(t)} \\
&\quad + \delta_2^{\bar{B}G} \bar{B}_2(t) \frac{G_2(t)}{T(t)} + \delta_2^{\bar{B}B} \bar{B}_2(t) \frac{B_2(t)}{T(t)}, \\
\bar{B}_2'(t) &= \alpha_2^B B_2(t) - \gamma_2^B \bar{B}_2(t) - \eta_2^B \bar{B}_2(t) \\
&\quad + \beta_2^{B\bar{G}} B_2(t) \frac{\bar{G}_2(t)}{T(t)} + \beta_2^{B\bar{B}} B_2(t) \frac{\bar{B}_2(t)}{T(t)} \\
&\quad - \delta_2^{\bar{B}G} \bar{B}_2(t) \frac{G_2(t)}{T(t)} - \delta_2^{\bar{B}B} \bar{B}_2(t) \frac{B_2(t)}{T(t)}. \\
T(t) &= G_1(t) + \bar{G}_1(t) + B_1(t) + \bar{B}_1(t) + G_2(t) + \bar{G}_2(t) + B_2(t) + \bar{B}_2(t)
\end{aligned} \tag{3}$$

The flow diagram, associated to the model, is plotted in Figure 2.

**Caption Figure 2**

Flow diagram of the epidemiological-mathematical model for dynamics of Bachillerato academic performance in Spain. The boxes represent the students depending on their sex, stage and academic results. The arrows denote the transits of students labeled by the expressions and parameters governing these transits.

### 2.3. Scaling the model

Data in Table 1 is in percentages meanwhile model (3) is referred to number of students. It leads us to transform (scaling) the model into the same units as data in order to fit the data with the model. To do that, we follow the ideas developed in [3, 3] about how to scale models where the population is varying in size. Here, we are not going to show the process and the scaled model because it is a technical transformation, the resulting equations are more complex and longer and does not provide extra information about the model. Moreover, the scaled model has the same parameters as the non-scaled model (3) with the same meaning. Now, in order to avoid introducing new notation, we are going to consider that the subpopulations  $G_1(t)$ ,  $\overline{G}_1(t)$ ,  $B_1(t)$ ,  $\overline{B}_1(t)$ ,  $G_2(t)$ ,  $\overline{G}_2(t)$ ,  $B_2(t)$ ,  $\overline{B}_2(t)$  correspond to the percentage of Girls and Boys in the promotable and non-promotable groups, in the First and Second course of *Bachillerato*.

## 3. Deterministic parameter estimation and prediction over the next few years

This section is addressed to estimate the parameters of model (3). This task has been performed by fitting the scaled model in the mean square sense to the available data collected in Table 1. Computations have been carried out with *Mathematica 8.0* [3]. The model (3) has been built using *epiModel* software [3]. *epiModel* facilitates the implementation all the equations in *Mathematica* saving developing time. Then, with a simple *Mathematica* algorithm, the model built by *epiModel* is scaled.

The system of differential equations (3), in its scaled version, is numerically solved by taking as initial conditions the data of the academic year 1999-2000 (corresponding to  $t = 0$ ), i.e., according to Table 1,  $G_1(0) = 19.68$ ,  $B_1(0) = 15.24$ ,  $\overline{G}_1(0) = 9.75$ ,  $\overline{B}_1(0) = 9.33$ ,  $G_2(0) = 16.21$ ,  $B_2(0) = 11.64$ ,

$\overline{G}_2(0) = 9.52$  and  $\overline{B}_2(0) = 8.63$ . Tables 2 and 3 collect the estimation of the model parameters.

In Table 2, we show these values for both, negative and positive autonomous decision and abandon rates. In Table 3, we show the corresponding values for negative and positive transmission rates.

### **Caption Table 2**

Estimation of positive and negative autonomous decision and abandon rates.

### **Caption Table 3**

Estimation of positive and negative transmission parameters.

Now, once the model is stated and the parameters estimated, we are able to give predictions of each subpopulation over the next few years by computing the solutions of the model for values of time  $t$  in the forthcoming future. In Table 4 we can see the model output for  $t = 1999 - 2000, \dots, 2014 - 2015$ . Note that from 2009-2010 the obtained values are the model predictions.

### **Caption Table 4**

The model output corresponding to the First and Second Stage of *Bachillerato*, in both, state and private high schools all over Spain during academic years 1999 – 2000 to 2014 – 2015. Each row shows the percentage of girls and boys who promote ( $G_i$  |  $B_i$ ) and do not promote ( $\overline{G}_i$  |  $\overline{B}_i$ ) for each level  $i = 1, 2$ . It can be compared the model output values for  $t = 1999 - 2000, \dots, 2008 - 2009$  to the data values in Table 1 to verify the goodness of the fitting. Graphically, it can be seen in the left-hand side of the graphs in Figure 3.

#### 4. Introducing uncertainty in the model parameters and predicting the next few years

Uncertainty is a key part of the real world and it should be considered in modelling, to be precise, in the data and the model parameters. Therefore, the assumption that parameters always are constant or the parameter estimation does not contain errors is not appropriate. Thus, it is natural to consider that the model parameters contain uncertainties. Hence, the deterministic prediction can give us an idea about the future trends but the obtained values may not be as accurate as expected.

Thus, we propose forecasting future evolutions using confidence intervals. In order to calculate these confidence intervals, let us use the technique called bootstrapping. Bootstrapping is a sophisticated and efficient method for determining a non-parametric probabilistic estimation of model parameters [3, 3, 3].

Specifically, the probabilistic estimation of the parameters is performed using a *residual* bootstrapping approach. In order to do it, and considering the general procedure published [3], we are going to follow the next steps:

- Step 1 Compute the error terms for the estimated parameters (deterministic parameters) and analyse and resample these error terms using bootstrapping.
- Step 2 Obtain new perturbed data by adding the resampled error (obtained in Step 1) to output data collected in Table 4 for  $t = 1999-2000, \dots, 2008-2009$ .
- Step 3 For each new data perturbation calculated (in Step 2), we compute the parameters that best fit the model (in the mean square sense).
- Step 4 For each set of parameter values obtained by fitting the model with the perturbed data, we solve the model with these parameters and compute the outputs in the required time instants.
- Step 5 Taking 95% confidence interval (of each output) from each subpopulation by percentile 2.5 and percentile 97.5 we will be able to conclude the percentage of students who promote/do not promote.

#### 4.1. Error term analysis

In order to analyse the error terms, we have followed the next steps:

- We compute the output of the model with the parameters in Tables 2 and 3 at the time instants  $t = 1999-2000, \dots, 2008-2009$  and compute their differences (errors) with the corresponding data from Table 1. Let us denote these errors by  $e_{G_1}(t)$ ,  $e_{G_2}(t)$ ,  $e_{B_1}(t)$ ,  $e_{B_2}(t)$ ,  $e_{\overline{G}_1}(t)$ ,  $e_{\overline{G}_2}(t)$ ,  $e_{\overline{B}_1}(t)$ ,  $e_{\overline{B}_2}(t)$  for each subpopulation. The results are shown in Table 5.

##### **Caption Table 5**

Differences between the real data in Table 1 and the output model in Table 4 corresponding to the First and Second Stage of *Bachillerato*, both state and private high schools all over Spain during academic years 1999 – 2000 to 2008 – 2009, for Girls|Boys who promote ( $G_i|B_i$ ) and do not promote ( $\overline{G}_i | \overline{B}_i$ ) for each level  $i = 1, 2$ .

- We analyse if the error terms in Table 5 are correlated. Pearson correlation coefficient was used. The results obtained indicate that the set of all pairs of errors were correlated (see Table 6).

##### **Caption Table 6**

Matrix of Pearson correlation coefficients for the error terms.

- Taking into account the Box-Ljung test [3], we also analyse if each error term is autocorrelated. Note that this non-parametric test can be used to check the hypothesis that the elements of a sequence are mutually independent. In this case, the results for girls in First Stage of *Bachillerato* are:  $p - value = 0.853$ ,  $p - value = 0.853$ ,  $p - value = 0.555$ ,  $p - value = 0.753$ ,  $p - value = 0.878$ ,  $p - value = 0.937$ ,  $p - value = 0.940$  and  $p - value = 0.965$ . None of the test statistic values is statistically significant ( $p - value > 0.05$ ), therefore the claim

that there is autocorrelation should be rejected. For the rest of the groups we obtain similar results, i.e., there are not autocorrelation in any of them.

- The normality of the distribution of errors is determined by using a non-parametric test. Taking into account that our error terms are correlated, *E-statistic (Energy) Test of Multivariate Normality* is applied [3]. In our case, this test has a p-value equal to 0.9963. Therefore, we can accept that errors  $e_{G_1}(t)$ ,  $e_{B_1}(t)$ ,  $e_{\bar{G}_1}(t)$ ,  $e_{\bar{B}_1}(t)$ ,  $e_{G_2}(t)$ ,  $e_{B_2}(t)$ ,  $e_{\bar{G}_2}(t)$  and  $e_{\bar{B}_2}(t)$  present a multivariate normal distribution. To be precise, we accept that

$$\mathbf{e}(t) = (e_{G_1}(t), e_{\bar{G}_1}(t), e_{B_1}(t), e_{\bar{B}_1}(t), e_{G_2}(t), e_{\bar{G}_2}(t), e_{B_2}(t), e_{\bar{B}_2}(t)),$$

where

$$\mathbf{e}(t) \sim N_8(\boldsymbol{\mu}_e, \boldsymbol{\Sigma}_e), \quad (4)$$

being the components of vector  $\boldsymbol{\mu}_e$  the expectations of each component of vector  $\mathbf{e}(t)$  and  $\boldsymbol{\Sigma}_e$  its variance-covariance matrix. These parameters have been estimated using the errors in Table 5:

$$\boldsymbol{\mu}_e = (0.0014, -0.0008, -0.0033, 0.0011, -0.0002, -0.0018, 0.0027, 0.0010),$$

$$\boldsymbol{\Sigma}_e = \begin{pmatrix} 1.3464 \cdot 10^{-4} & \dots & -4.5276 \cdot 10^{-5} \\ 2.8025 \cdot 10^{-5} & \dots & -1.1065 \cdot 10^{-5} \\ \vdots & \ddots & \vdots \\ -4.52763 \cdot 10^{-5} & \dots & 1.9350 \cdot 10^{-5} \end{pmatrix}.$$

#### 4.2. Generating new output perturbed data

The process carried out to generate new output perturbed data has been the following. For each one of the 1000 random error terms following the multivariate normal distribution given by the expression (4) assumed by *E-statistic (Energy) Test of Multivariate Normality*:

- We add these error terms (1000 times) to data in Table 4 for  $t = 1999 - 2000, \dots, 2008 - 2009$ , obtaining a new set of perturbed data. Note that we obtained 1000 sets of output perturbed data.

- And we compute the parameters which best fit the model with the set of perturbed data (in the least square sense) and store them, using the same procedure we used to estimate the parameters in Tables 2 and 3. Note that this procedure allows us to have 1000 sets of values for the parameters of the model.

#### 4.3. Obtaining confidence intervals for model outputs

Finally, the confidence intervals are obtained as follows:

- For each one of the 1000 set of parameters, we solve the system of differential equations (3) in order to compute the model output for each subpopulation of students and  $t = 2009 - 2010, \dots, 2014 - 2015$ .
- For each  $t$  and each subpopulation, we have a set of 1000 model output values. Then, we compute the mean, median and the 95% confidence interval by percentiles 2.5 and 97.5. These confidence intervals give us the non-parametric probabilistic prediction of the evolution in the next few years. The obtained results can be seen in Table 7.

#### Caption Table 7

The 95% confidence interval prediction corresponding to the First and Second Stage of *Bachillerato*, in both, state and private high schools all over Spain during academic years 2009 – 2010 to 2014 – 2015. Each row shows the percentage of girls/boys who promote ( $G_i/B_i$ ) and do not promote ( $\overline{G}_i/\overline{B}_i$ ) for each level  $i = 1, 2$ .

Thus, in Figure 3 we can see graphically, for each subpopulation, the data from Table 4 (red points), the deterministic model prediction (line), the 95% confidence intervals (error bars). The points in the middle of the confidence intervals are the median of the 1000 outputs for each subpopulation at each time instant where we have data about the academic results of Spanish *Bachillerato* students. These median values are the ones appearing in Table 7. Also, we can observe that there is a slight decreasing in the non-promotable groups. However, the sum of the students in non-promotable

groups is around 30% which is unacceptable in a modern society. Furthermore, we can see how uncertainty increases in the predictions for the two Stages of promoted girls and the First Stage of promoted boys.

### Caption Figure 3

Real data (red points on the left side of vertical axis) and prediction (line) with confidence intervals (on the right side of vertical axis) of the academic performance of *Bachillerato* Spanish students over the academic years 1999–2000 to 2014–2015. Smaller confidence intervals, represent less uncertainty in the predictions, the points in the middle of the confidence intervals are their medians. The square black point represents the last academic results published recently corresponding to the academic year 2009–2010. A good prediction is done if the black point lies inside the 95% confidence interval. Notice that each graph has its own scale.

As time goes, new academic results have been published, in particular, the academic results of students of *Bachillerato* in the academic year 2009–2010, data that could not be used initially to fit the model because they were not available at that time. These new data allow us to compare the obtained predictions from our model with the new real data.

In Figure 3, we also include, for each subpopulation, the obtained academic results during the academic year 2009–2010 (square black points). These values are the ones appearing in Table 10. Except for the non-promoted students of Second Stage of *Bachillerato*, we can see that probabilistic predictions by confidence intervals provide good results.

We can see that probabilistic predictions by 95% confidence intervals provide good results (the black points lie inside the confidence intervals) for all the Stages except the non-promotable groups in the Second Stage of *Bachillerato* where the distances between the black points and the lines (model predictions) are in each graph around 1% of the total students.

This 1% decreasing of non-promoted students in the Second Stage of *Bachillerato* seems to come from academic courses 2007–2008. This may be explained by:



- According to [3], the perception of the economical crisis motivates the improvement of the student qualification (future workers), especially, in the Second Stage of Bachillerato because at the end, students obtain the degree.
- Furthermore, in the academic course 2009 – 2010 the Spanish Education law passed in February 2010 [4] to adapt the university access to Bologna Plan permits the students in the course 2009-2010 had less restrictions to choose the subjects in this academic level. It made that the students could choose subjects traditionally with higher success rates (Earth Sciences rather than Physics or avoiding Mathematics in Social Sciences branch) to achieve a better academic results to access the university.

However, we consulted with educational experts and they consider that the decreasing of the non-promotable groups in Second Stage of Bachillerato (around 2% of the total Bachillerato students) is not significative.

### Caption Table 10

The 95% confidence interval prediction and the real academic results corresponding to the First and Second Stage of *Bachillerato*, in both, state and private high schools all over Spain during the academic year 2009 – 2010. Each row shows the rate of girls/boys who promoted ( $G_i/B_i$ ) and did not promote ( $\overline{G}_i/\overline{B}_i$ ) for each academic level  $i = 1, 2$ .

## 5. Abandon analysis

One of the most difficult aspects in academic performance is the study and analysis of the abandon, because there are not much available data and the experts still do not agree with the right definition [4]. In fact, we have made a decision in order to include this issue in the model and this is to consider *abandon* when, during the academic year, the student leaves the academic system. The student may resume her/his studies in the future, but our model will consider her/him as a new student. Thus, the model allows us to quantify the number of students who leave yearly the system by computing

$$\int_t^{t+1} (\eta_1^G \overline{G}_1(s) + \eta_2^G \overline{G}_2(s) + \eta_1^B \overline{B}_1(s) + \eta_2^B \overline{B}_2(s)) ds,$$

where  $t = 2009 - 2010, \dots, 2013 - 2014$ . The results are collected in Table 8.

### **Caption Table 8**

Estimation of the percentage of abandon in Spanish *Bachillerato* during the academic years from 1999–2000 to 2014–2015.

Moreover, taking advantage of the bootstrapping analysis carried out in the previous section, we can predict the evolution of abandon rate in the next few years by means of 95% confidence intervals. The obtained results are shown in Table 9.

### **Caption Table 9**

Descriptive analysis of the percentage of abandon in Spanish *Bachillerato* during the academic years from 2009–2010 to 2014–2015.

In both tables we can see a slight decreasing in the number of students who leave the high school.

## **6. Conclusion**

In this paper we propose a dynamical model to study the students' academic performance in high school in Spain, taking into account sex, stages and academic results. The main idea behind our approach is to consider that academic performance depends on both student own study and their classmates' habits. The abandon is a crucial issue to analyse school failure that has also been contemplated in the model. To make more realistic our approach, we have considered uncertainty in the study. This fact allows us to predict the students' academic performance in the next few years through confidence intervals. The model predictions for 2009 – 2010 have been compared with the recently available data with good predictions for most of the student groups.

The results tell us that there is a slight decreasing of the number of students in the non-promotable groups and who leave the high school, and it seems to reach a stationary situation. The current and predicted scenarios are very worrying because around 30% of the students have bad academic results.

We would like to point out that we have built a model using few data with respect to the number of model parameters and including hypotheses appearing in the literature that rarely are quantified. In spite of this, we have obtained good predictions for most of the student groups. We also would notice that one of the major inconveniences we have found during this research is the obtention of reliable data, because the Administration is reluctant to facilitate these kind of data due to it is, in Spain, a sensitive political concern. In our opinion, the availability of a good data set, we are still working on, is crucial to disaggregate the factors involved in each model parameter and quantify them.

Finally, we would like to say that the modelling technique presented in this paper can be used in other educational levels in any region or country.

- [1] Ley Orgánica 1/1990, de 3 de octubre, de Ordenación General del Sistema Educativo. Boletín Oficial de Estado de 4 de octubre de 1990. [Organic Law 1/1990 of October 3, on the General Educational System. Official State Bulletin on October 4, 1990]. (In Spanish). Available at [http://www.boe.es/aeboe/consultas/bases/\\_datos/doc.php?id=BOE-A-1990-24172](http://www.boe.es/aeboe/consultas/bases/_datos/doc.php?id=BOE-A-1990-24172) (Accessed on April 24, 2012).
- [2] Ley Orgánica 2/2006, de 3 de mayo, de Educación. Boletín Oficial de Estado, 106, de 4 de mayo de 2006, 17158–17207. [Organic Law 2/2006 of May 3, on Education. Official State Bulletin, 106, of May 4, 2006, 17158–17207]. (In Spanish). Available at [http://www.boe.es/aeboe/consultas/bases/\\_datos/doc.php?id=BOE-A-2006-7899](http://www.boe.es/aeboe/consultas/bases/_datos/doc.php?id=BOE-A-2006-7899) (Accessed on April 24, 2012).
- [3] A. Bolívar Botía, L. López Calvo, Las grandes cifras del fracaso y los riesgos de exclusión educativa. Profesorado, Revista de Currículum y Formación de Profesorado. 13(3) (2009). [The large numbers of academic underachievement and the risk of educational exclusion]. (In Spanish). Available at <http://www.ugr.es/~recfpro/rev133ART2.pdf> (Accessed on April 24, 2012).
- [4] A. Marchesi, C. Hernández, El Fracaso Escolar. Perspectiva Internacional, Alianza, Madrid, 2003. [Academic Underachievement. International Perspective]. (In Spanish). ISBN:84-206-2955-3.
- [5] <http://www.rlp.com.ni/noticias/93005> (Accessed on April 24, 2012).
- [6] <http://www.euractiv.com/en/enterprise-jobs/eu-youth-job-strategy-under-fire-news-497858> (Accessed on April 24, 2012).
- [7] J. Casal, M. García, Las reformas en los dispositivos de formación para combatir el fracaso escolar en Europa: paradojas de un éxito, Revista de Educación. 317 (1998) 301-317. [Reforms in training measures to combat academic underachievement in Europe: paradoxes of success]. (In Spanish). Available at <http://www.doredin.mec.es/documentos/00820073004064.pdf> (Accessed on April 24, 2012).

- [8] H. Eckert, Entre el fracaso escolar y las dificultades de inserción profesional: la vulnerabilidad de los jóvenes sin formación en el inicio de la sociedad del conocimiento, *Revista de Educación*. 341 (2006) 35-55. [Between academic underachievement and employability difficulties: the vulnerability of young people without training in the beginning of the knowledge society]. (In Spanish). Available at [http://www.oei.es/etp/fracaso\\\_escolar\\\_dificultades\\\_insercion\\\_profesional\\\_eckert.pdf](http://www.oei.es/etp/fracaso\_escolar\_dificultades\_insercion\_profesional\_eckert.pdf) (Accessed on April 24, 2012).
- [9] A. Marchesi, R. Lucena, *La Representación Social del Fracaso Escolar*, Alianza, Madrid, 2003. [The Social Representation of Academic Underachievement]. (In Spanish). Available at [http://www.ticnologia.es/contenidos/sociologia/T11\\\_1.pdf](http://www.ticnologia.es/contenidos/sociologia/T11\_1.pdf) (Accessed on April 24, 2012).
- [10] Real Decreto 1467/2007, de 2 de noviembre por el que se establece la estructura del *Bachillerato* español y se fijan sus enseñanzas mínimas. [Royal Decree 1467/2007 of 2 November for establishing the structure of Spanish high school and set their minimum contents] (In Spanish). Available at [http://www.boe.es/aeboe/consultas/bases\\\_datos/doc.php?id=BOE-A-2007-19184](http://www.boe.es/aeboe/consultas/bases\_datos/doc.php?id=BOE-A-2007-19184) (Accessed on April 24, 2012).
- [11] J.I. Pozo, *Teorías Cognitivas del Aprendizaje*, Morata, Madrid, 1989. [Cognitive Theories of Learning]. (In Spanish). ISBN-10:84-7112-335-5.
- [12] A. Tryphon, J. Vonche, *Piaget-Vygotsky: La Génesis Social del Pensamiento*, Paidós Educador, México, 2000. [Piaget-Vygotsky: The Social Genesis of Thought]. (In Spanish). ISBN:9789501221503.
- [13] L.S. Vygotsky, *Mind in Society: The Development of Higher Mental Processes*, Harvard University Press, Cambridge, 1978. ISBN-10:0674576292.
- [14] M. Lucci, La propuesta de Vygotsky: La psicología socio-histórica, *Revista de Currículo y Formación del Profesorado*. 10(2) (2006) 7-11. [The proposal of Vygotsky: socio-historical psychology]. (In Spanish). Available at <http://www.ugr.es/~recfpro/rev102C0L2.pdf> (Accessed on April 24, 2012).
- [15] H. Daniels, M. Cole, J. Wertsch, *Cambridge Companion to Vygotski*, Cambridge University Press, Cambridge, 2007. ISBN-10:0521537878.

- [16] N.A. Christakis, J.H. Fowler, *Connected: The Surprising Power Of Our Social Networks And How They Shape Our Lives*, Hachette Book Group, Brown and Company, 2009. ISBN-10:0316036145.
- [17] N.A. Christakis, J.H. Fowler, The spread of obesity in a large social network over 32 years, *The New England Journal of Medicine*. 357 (2007) 370-379. DOI:10.1056/NEJMsa066082.
- [18] F.J. Santonja, A. Morales, R.J. Villanueva, J.C. Cortés, Analysing the effect of public health campaigns on reducing excess weight: A modelling approach for the Spanish Autonomous Region of the Community of Valencia, *Evaluation and Program Planning*. 35(1) (2012) 34-39. DOI:10.1016/j.evalprogplan.2011.06.004.
- [19] F.J. Santonja, E. Sánchez, M. Rubio, M, J. Morera, J, Alcohol consumption in Spain and its economic cost: A mathematical modelling approach, *Mathematical and Computer Modelling*. 52(7-8) (2010) 999-1003. DOI:10.1016/j.mcm.2010.02.029.
- [20] E. Sánchez, R.J. Villanueva, F.J. Santonja, M. Rubio, Predicting cocaine consumption in Spain: A mathematical modelling approach, *Drugs: Education, Prevention and Policy*. 18(2) (2011) 108-115. DOI:10.3109/09687630903443299.
- [21] I. García, L. Jódar, P. Merello, F.J. Santonja, A discrete mathematical model for addictive buying: predicting the affected population evolution, *Mathematical and Computer Modelling*. 54(78) (2011) 1634-1637. DOI:10.1016/j.mcm.2010.12.012.
- [22] L.M.A. Bettencourt, A. Cintron-Arias, D.I. Kaiser, C. Castillo-Chavez, The power of a good idea: Quantitative modelling of the spread of ideas from epidemiological models, *Physica A*. 364 (2006) 513-536. DOI:10.1016/j.physa.2005.08.083.
- [23] M. Peco, F.J. Santonja, A.C. Tarazona, R.J. Villanueva, J. Villanueva-Oller, The effect of the Spanish Law of Political Parties (LPP) on the attitude of the Basque Country population towards ETA: A dynamic modelling approach, *Mathematical and Computer Modelling*. (2011). DOI:10.1016/j.mcm.2011.11.007.

- [24] J. Camacho, J.C. Cortés, R.M. Micle, A. Sánchez-Sánchez, Predicting the academic underachievement in high school in Spain over the next few years: A dynamic modelling approach, *Mathematical and Computer Modelling*. (2011). DOI: 10.1016/j.mcm.2011.11.011. (To be published).
- [25] T. Susinos, Un recorrido por la inclusión educativa española, *Investigaciones y experiencias más recientes*, *Revista de Educación*. 327 (2002) 49-56 A. [A tour of the Spanish educational inclusion]. (In Spanish). Available at <http://www.doredin.mec.es/documentos/008200330016.pdf> (Accessed on April 24, 2012).
- [26] <http://www.educacion.gob.es/horizontales/estadisticas/no-universitaria/alumnado/resultados.html> (Accessed on April 24, 2012).
- [27] R. Wentzel Kathryn, D. Watkins, Peer relationships and collaborative learning as contexts for academic enablers, *School Psychology Review*. 31(3) (2002) 366-377. Available at <http://www.bibsonomy.org/bibtex/2402c8d2eb020c0aa3f46f9f9ca7d69cb/clachapelle> (Accessed on April 24, 2012).
- [28] J.D. Murray, *Mathematical Biology*, Springer, 2002. ISBN-10:0387952233.
- [29] C. Fierro-Hernández, Patrón de rasgos personales y comportamiento escolar en jóvenes, *Revista de Educación*. 332 (2000) 291-304. [Personal characteristic patterns and scholar behavior in your people]. (In Spanish). ISSN:0034-8082.
- [30] M. Martcheva, C. Castillo-Chavez, Diseases with chronic stage in a population with varying size, *Mathematical Biosciences*. 182(1) (2003) 1-25. DOI:10.1016/S0025-5564(02)00184-0.
- [31] J. Mena-Lorca, H.W. Hethcote, Dynamic models of infectious diseases as regulators of population sizes, *Journal of Mathematical Biology*. 30 (1992) 693-716. DOI:10.1007/BF00173264.
- [32] <http://www.wolfram.com/products/mathematica>. (Accessed on April 24, 2012).

- [33] J.C. Cortés, A. Sánchez-Sánchez, F.J. Santonja, R.J. Villanueva, *epiModel*: A system to build automatically systems of differential equations of compartmental type-epidemiological models, *Computers in Biology and Medicine*. 41(11) (2011) 999-1005. DOI:10.1016/j.complbiomed.2011.06.018.
- [34] G. Dogan, Bootstrapping for confidence interval estimation and hypothesis for parameters of system dynamics models, *System Dynamics Review*. 23 (2007) 415-436. DOI:10.1002/sdr.362.
- [35] M. Joshi, A. Seidel-Morgenstern, A. Kremling, Exploiting the bootstrap method for quantifying parameter confidence intervals in dynamical systems, *Metabolic Engineering*. 8(5) (2006) 447-455. DOI:10.1016/j.ymben.2006.04.003.
- [36] W.C.M. Van Beers, J.P.C. Kleijnen, Customized sequential designs for random simulation experiments: Kriging metamodeling and bootstrapping, *European Journal of Operational Research*. 186(3) (2008) 1099-1113. DOI:10.1016/j.ejor.2007.02.035.
- [37] G. Ljung, G. Box, On a mesasure of lack of fit in time series models, *Biometrika*. 66 (1979) 265-270. Available at <http://www.jstor.org/stable/2335207> (Accessed on April 24, 2012).
- [38] G.J. Szekely, M.L. Rizzo, A new test for multivariate normality, *Journal of Multivariate Analysis*. 93(1) (2005) 58-80. DOI:10.1016/j.jmva.2003.12.002.
- [39] A. Moreno Mínguez, A. López Peláez, S. Segado Sánchez-Cabezudo, *The Transtition to Adulthood in Spain. Economic Crisis and Late Emancipation*, Social Studies Collection, 'La Caixa' Welfare Projects, Barcelona, 2012. ISBN: 978-84-9900-070-1. Available at [http://obrasocial.lacaixa.es/StaticFiles/StaticFiles/85143c805cc68310VgnVCM200000128cf10aRCRD/es/vol134\\_en.pdf](http://obrasocial.lacaixa.es/StaticFiles/StaticFiles/85143c805cc68310VgnVCM200000128cf10aRCRD/es/vol134_en.pdf) (Accessed on October 15, 2012).
- [40] Orden EDU/268/2010, de 11 de febrero, por la que se modifica la Orden EDU/1434/2009, de 29 de mayo, por la que se actualizan los anexos del Real Decreto 1892/2008, de 14 de noviembre, por el que se regulan las condiciones para el acceso a las enseñanzas universitarias oficiales de grado y los procedimientos de admisión a las universidades



públicas españolas. [EDU/268/2010 Order of 11 February, amending EDU/1434/2009 Order of 29 May, which update the annexes of Royal Decree 1892/2008 of 14 November, regulating the conditions for access to official university degree and admission procedures to the Spanish public universities.] (In Spanish). Available at <http://www.boe.es/boe/dias/2010/02/13/pdfs/BOE-A-2010-2330.pdf> (Accessed on October 15, 2012).

- [41] J. Calero, A. Choi, S. Waisgrais, Determinantes del riesgo de fracaso escolar en España: una aproximación a través de un análisis logístico multinivel aplicado a PISA-2006, *Revista de Educación*, número extraordinario 2010: Abandono temprano de la educación y la formación: cifras y políticas. (2010) 225–256. [Risk factors of academic underachievement in Spain: an approach through a multilevel logistic analysis applied to PISA-2006]. (In Spanish). Retrieved April 24, 2012, from: [http://www.revistaeducacion.mec.es/re2010/re2010\\\_09.pdf](http://www.revistaeducacion.mec.es/re2010/re2010\_09.pdf)

Academic year	First Stage (Girls   Boys)		Second Stage (Girls   Boys)	
	% Promote ( $G_1$   $B_1$ )	% Non-Promote ( $\bar{G}_1$   $\bar{B}_1$ )	% Promote ( $G_2$   $B_2$ )	% Non-Promote ( $\bar{G}_2$   $\bar{B}_2$ )
1999–2000	19.68   15.24	9.75   9.33	16.21   11.64	9.52   8.63
2000–2001	22.65   17.54	9.91   10.12	14.07   10.04	8.24   7.43
2001–2002	19.23   14.23	8.61   9.10	17.86   13.06	9.32   8.59
2002–2003	18.87   14.19	8.36   8.51	19.14   13.97	8.76   8.20
2003–2004	19.93   15.06	7.74   7.88	19.19   13.80	8.44   7.96
2004–2005	20.11   15.14	7.65   7.94	18.90   13.92	8.39   7.95
2005–2006	20.07   15.39	7.64   7.93	19.14   13.97	8.08   7.78
2006–2007	20.06   15.34	7.67   7.87	19.14   14.29	7.98   7.65
2007–2008	20.25   15.82	7.57   7.66	19.37   14.61	7.60   7.12
2008–2009	20.72   16.57	7.28   7.43	19.43   14.86	7.05   6.66

Table 1: The available data corresponding to the First and Second Stage of Bachillerato, in both, state and private high schools all over Spain from academic year 1999 – 2000 to 2008 – 2009. Each row shows the percentage of Girls and Boys who promote ( $G_i$  |  $B_i$ ) and do not promote ( $\bar{G}_i$  |  $\bar{B}_i$ ) for each level  $i = 1, 2$  over the total Spanish *Bachillerato* students.

Gender	Negative autonomous decision		Positive autonomous decision		Abandon	
	Parameter	Value	Parameter	Value	Parameter	Value
Girls	$\alpha_1^G$	0.04501	$\gamma_1^G$	0.08685	$\eta_1^G$	0.07480
	$\alpha_2^G$	0.00366	$\gamma_2^G$	0.00385	$\eta_2^G$	0.06431
Boys	$\alpha_1^B$	0.04610	$\gamma_1^B$	0.11643	$\eta_1^B$	0.02676
	$\alpha_2^B$	0.01208	$\gamma_2^B$	0.04163	$\eta_2^B$	0.00232

Table 2: Estimation of positive and negative autonomous decision and abandon rates.

Gender	Negative transmission		Positive transmission	
	Parameter	Value	Parameter	Value
Girls	$\beta_1^{G\bar{G}}$	0.00002	$\delta_1^{\bar{G}G}$	0.03699
	$\beta_1^{G\bar{B}}$	0.11093	$\delta_1^{\bar{G}B}$	0.09793
	$\beta_2^{G\bar{G}}$	0.08939	$\delta_2^{\bar{G}G}$	0.00607
	$\beta_2^{G\bar{B}}$	0.09837	$\delta_2^{\bar{G}B}$	0.06962
Boys	$\beta_1^{B\bar{G}}$	0.08700	$\delta_1^{\bar{B}G}$	0.01881
	$\beta_1^{B\bar{B}}$	0.02852	$\delta_1^{\bar{B}B}$	0.04922
	$\beta_2^{B\bar{G}}$	0.01837	$\delta_2^{\bar{B}G}$	0.11703
	$\beta_2^{B\bar{B}}$	0.11679	$\delta_2^{\bar{B}B}$	0.07805

Table 3: Estimation of positive and negative transmission parameters.

Academic year	First Stage of <i>Bachillerato</i> (Girls   Boys)		Second Stage of <i>Bachillerato</i> (Girls   Boys)	
	% Promote ( $G_1$   $B_1$ )	% Non-Promote ( $\bar{G}_1$   $\bar{B}_1$ )	% Promote ( $G_2$   $B_2$ )	% Non-Promote ( $\bar{G}_2$   $\bar{B}_2$ )
1999–2000	19.68   15.24	9.75   9.33	16.21   11.64	9.52   8.63
2000–2001	19.94   15.29	9.24   9.07	16.73   12.13	9.16   8.44
2001–2002	20.14   15.34	8.80   8.82	17.18   12.61	8.85   8.26
2002–2003	20.29   15.38	8.43   8.56	17.58   13.07	8.60   8.09
2003–2004	20.39   15.42	8.11   8.31	17.94   13.51	8.39   7.93
2004–2005	20.46   15.46	7.84   8.07	18.25   13.93	8.21   7.78
2005–2006	20.49   15.50	7.60   7.83	18.54   14.35	8.06   7.63
2006–2007	20.49   15.53	7.40   7.60	18.79   14.75	7.94   7.50
2007–2008	20.47   15.57	7.22   7.37	19.02   15.14	7.84   7.37
2008–2009	20.44   15.60	7.06   7.15	19.22   15.51	7.76   7.26
2009–2010	20.38   15.63	6.92   6.94	19.40   15.88	7.69   7.15
2010–2011	20.31   15.67	6.80   6.73	19.57   16.43	7.64   7.05
2011–2012	20.22   15.70	6.69   6.53	19.72   16.58	7.60   6.96
2012–2013	20.13   15.72	6.59   6.34	19.86   16.92	7.56   6.87
2013–2014	20.03   15.75	6.50   6.16	19.99   17.25	7.54   6.79
2014–2015	19.91   15.78	6.42   5.98	20.10   17.57	7.52   6.72

Table 4: The model output corresponding to the First and Second Stage of *Bachillerato*, in both, state and private high schools all over Spain during academic years 1999 – 2000 to 2014 – 2015. Each row shows the percentage of girls and boys who promote ( $G_i$  |  $B_i$ ) and do not promote ( $\bar{G}_i$  |  $\bar{B}_i$ ) for each level  $i = 1, 2$ . It can be compared the model output values for  $t = 1999 - 2000, \dots, 2008 - 2009$  to the data values in Table 1 to verify the goodness of the fitting. Graphically, it can be seen in the left-hand side of the graphs in Figure 3.

Academic year	First Stage of <i>Bachillerato</i> (Girls   Boys)		Second Stage of <i>Bachillerato</i> (Girls   Boys)	
	% Promote $e_{G_1}(t)$   $e_{B_1}(t)$	% Non-Promote $e_{\bar{G}_1}(t)$   $e_{\bar{B}_1}(t)$	% Promote $e_{G_2}(t)$   $e_{B_2}(t)$	% Non-Promote $e_{\bar{G}_2}(t)$   $e_{\bar{B}_2}(t)$
2000–2001 (t=1)	-0.02699   -0.02250	-0.00660   -0.01048	0.02657   0.02091	0.00919   0.01009
2001–2002 (t=2)	0.00901   0.01107	0.00194   -0.00293	-0.00678   -0.00454	-0.00465   -0.00331
2002–2003 (t=3)	0.01418   0.01191	0.00072   0.00054	-0.01556   -0.00907	-0.00160   -0.00112
2003–2004 (t=4)	0.00462   0.00372	0.00373   0.00435	-0.01251   -0.00294	-0.00053   -0.00033
2004–2005 (t=5)	0.00347   0.00321	0.00189   0.00131	-0.00645   0.00013	-0.00181   -0.00175
2005–2006 (t=6)	0.00419   0.00099	-0.00038   -0.00097	-0.00604   0.00377	-0.00019   -0.00147
2006–2007 (t=7)	0.00434   0.00195	-0.00273   -0.00270	-0.00353   0.00458	-0.00041   -0.00151
2007–2008 (t=8)	0.00225   -0.00241	-0.00351   -0.00286	-0.00357   0.00526	0.00239   0.00255
2008–2009 (t=9)	-0.00284   -0.00968	-0.00217   -0.00276	-0.00204   0.00653	0.00707   0.00598

Table 5: Differences between the real data in Table 1 and the output model in Table 4 corresponding to the First and Second Stage of *Bachillerato*, both state and private high schools all over Spain during academic years 1999 – 2000 to 2008 – 2009, for Girls|Boys who promote ( $G_i|B_i$ ) and do not promote ( $\bar{G}_i$  |  $\bar{B}_i$ ) for each level  $i = 1, 2$ .

		$e_{G_1}(t)$	$e_{B_1}(t)$	$e_{\overline{G}_1}(t)$	$e_{\overline{B}_1}(t)$	$e_{G_2}(t)$	$e_{B_2}(t)$	$e_{\overline{G}_2}(t)$	$e_{\overline{B}_2}(t)$
$e_{G_1}(t)$	Pearson correlation coefficient	1	0.744	-0.968	-0.851	0.957	0.783	-0.945	-0.887
	P-Value		0.022*	0.000*	0.004*	0.000*	0.012*	0.000*	0.001*
$e_{B_1}(t)$	Pearson correlation coefficient	0.744	1	-0.811	-0.791	0.803	0.880	-0.860	-0.775
	P-Value	0.022*		0.008*	0.011*	0.009*	0.002*	0.003*	0.014*
$e_{\overline{G}_1}(t)$	Pearson correlation coefficient	-0.968	-0.811	1	0.771	-0.897	-0.900	0.939	0.819
	P-Value	0.000*	0.008*		0.015*	0.001*	0.001*	0.000*	0.007*
$e_{\overline{B}_1}(t)$	Pearson correlation coefficient	-0.851	-0.791	0.771	1	-0.951	-0.661	0.851	0.978
	P-Value	0.004*	0.011*	0.015*		0.000*	0.053*	0.004*	0.000*
$e_{G_2}(t)$	Pearson correlation coefficient	0.957	0.803	-0.897	-0.951	1	0.735	-0.961	-0.940
	P-Value	0.000*	0.009*	0.001*	0.000*		0.024*	0.000*	0.000*
$e_{B_2}(t)$	Pearson correlation coefficient	0.783	0.880	-0.900	-0.661	0.735	1	-0.813	-0.710
	P-Value	0.012*	0.002*	0.001*	0.053*	0.024*		0.008*	0.032*
$e_{\overline{G}_2}(t)$	Pearson correlation coefficient	-0.945	-0.860	0.939	0.851	-0.961	-0.813	1	0.836
	P-Value	0.000*	0.003*	0.000*	0.004*	0.000*	0.008*		0.005*
$e_{\overline{B}_2}(t)$	Pearson correlation coefficient	-,887	-,775	,819	,978	-,940	-,710	0.836	1
	P-Value	0.001*	0.014*	0.007*	0.000*	0.000*	0.032*	0.005*	

\* P-Value statistically significant to 95%

Table 6: Matrix of Pearson correlation coefficients for the error terms.

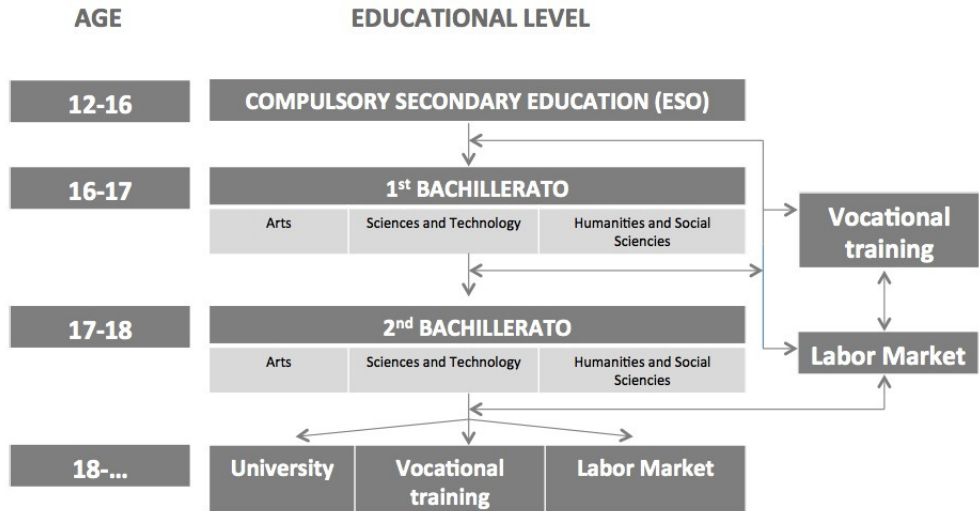


Figure 1:

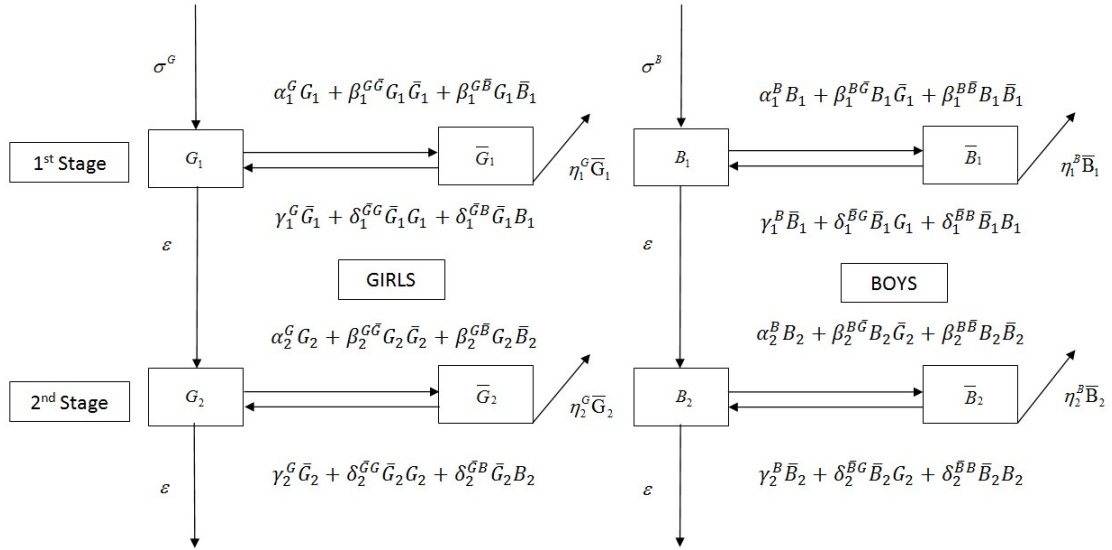


Figure 2:

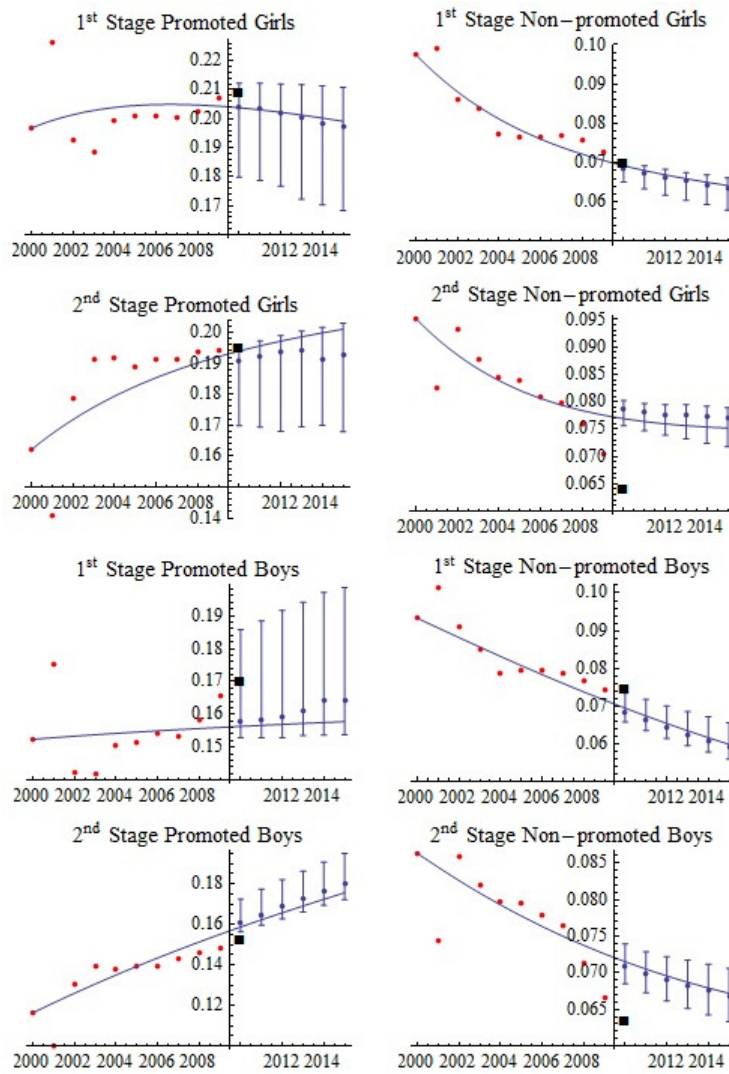


Figure 3:

**Caption Figure 1**

Structure of the Spanish educational system for students aged 12-18. Most of the students follow the academic path ESO + *Bachillerato*.

**Caption Figure 2**

Flow diagram of the epidemiological-mathematical model for dynamics of *Bachillerato* academic performance in Spain. The boxes represent the students depending on their sex, stage and academic results. The arrows denote the transits of students labeled by the expressions and parameters governing these transits.

**Caption Figure 3**

Real data (red points on the left side of vertical axis) and prediction (line) with confidence intervals (on the right side of vertical axis) of the academic performance of *Bachillerato* Spanish students over the academic years 1999–2000 to 2014–2015. Smaller confidence intervals, represent less uncertainty in the predictions, the points in the middle of the confidence intervals are their medians. The square black point represents the last academic results published recently corresponding to the academic year 2009 – 2010. A good prediction is done if the black point lies inside the confidence interval. Notice that each graph has its own scale.



Group	Time (t)	Mean	Median	Confidence interval
$G_1$	2010	0.20238	0.20414	[ 0.17993 , 0.21227 ]
$\overline{G}_1$	2010	0.06855	0.06866	[ 0.06512 , 0.07041 ]
$B_1$	2010	0.18888	0.19092	[ 0.16987 , 0.19554 ]
$\overline{B}_1$	2010	0.07850	0.07875	[ 0.07564 , 0.08020 ]
$G_2$	2010	0.16059	0.15807	[ 0.15286 , 0.18575 ]
$\overline{G}_2$	2010	0.06848	0.06834	[ 0.06574 , 0.07340 ]
$B_2$	2010	0.16161	0.16099	[ 0.15632 , 0.17254 ]
$\overline{B}_2$	2010	0.07099	0.07097	[ 0.06852 , 0.07391 ]
$G_1$	2011	0.20158	0.20346	[ 0.17870 , 0.21222 ]
$\overline{G}_1$	2011	0.06723	0.06734	[ 0.06333 , 0.06920 ]
$B_1$	2011	0.19030	0.19241	[ 0.16938 , 0.19734 ]
$\overline{B}_1$	2011	0.07791	0.07816	[ 0.07471 , 0.07969 ]
$G_2$	2011	0.16121	0.15852	[ 0.15283 , 0.18840 ]
$\overline{G}_2$	2011	0.06640	0.06624	[ 0.06346 , 0.07173 ]
$B_2$	2011	0.16542	0.16470	[ 0.15965 , 0.17752 ]
$\overline{B}_2$	2011	0.06994	0.06991	[ 0.06728 , 0.07287 ]
$G_1$	2012	0.19997	0.20220	[ 0.17673 , 0.21202 ]
$\overline{G}_1$	2012	0.06612	0.06630	[ 0.06179 , 0.06830 ]
$B_1$	2012	0.19085	0.19355	[ 0.16803 , 0.19898 ]
$\overline{B}_1$	2012	0.07748	0.07770	[ 0.07392 , 0.07949 ]
$G_2$	2012	0.16261	0.15924	[ 0.15291 , 0.19165 ]
$\overline{G}_2$	2012	0.06447	0.06430	[ 0.06137 , 0.07001 ]
$B_2$	2012	0.16942	0.16870	[ 0.16287 , 0.18219 ]
$\overline{B}_2$	2012	0.06902	0.06899	[ 0.06613 , 0.07221 ]
$G_1$	2013	0.19750	0.20043	[ 0.17223 , 0.21172 ]
$\overline{G}_1$	2013	0.06515	0.06536	[ 0.06044 , 0.06749 ]
$B_1$	2013	0.19047	0.19409	[ 0.16952 , 0.20041 ]
$\overline{B}_1$	2013	0.07719	0.07760	[ 0.07322 , 0.07948 ]
$G_2$	2013	0.16490	0.16126	[ 0.15361 , 0.19406 ]
$\overline{G}_2$	2013	0.06271	0.06247	[ 0.05941 , 0.06848 ]
$B_2$	2013	0.17365	0.17278	[ 0.16616 , 0.18630 ]
$\overline{B}_2$	2013	0.06826	0.06826	[ 0.06513 , 0.07170 ]
$G_1$	2014	0.19516	0.19850	[ 0.17036 , 0.21135 ]
$\overline{G}_1$	2014	0.06421	0.06444	[ 0.05934 , 0.06691 ]
$B_1$	2014	0.19019	0.19112	[ 0.16989 , 0.20172 ]
$\overline{B}_1$	2014	0.07691	0.07736	[ 0.07242 , 0.07916 ]
$G_2$	2014	0.16697	0.16410	[ 0.15372 , 0.19705 ]
$\overline{G}_2$	2014	0.06102	0.06072	[ 0.05772 , 0.06713 ]
$B_2$	2014	0.17770	0.17656	[ 0.16943 , 0.19088 ]
$\overline{B}_2$	2014	0.06756	0.06760	[ 0.06426 , 0.07116 ]
$G_1$	2015	0.19383	0.19730	[ 0.16837 , 0.21089 ]
$\overline{G}_1$	2015	0.06321	0.06351	[ 0.05798 , 0.06610 ]
$B_1$	2015	0.19100	0.19262	[ 0.16786 , 0.20304 ]
$\overline{B}_1$	2015	0.07650	0.07694	[ 0.07180 , 0.07888 ]
$G_2$	2015	0.16782	0.16442	[ 0.15377 , 0.19860 ]
$\overline{G}_2$	2015	0.05933	0.05898	[ 0.05586 , 0.06554 ]
$B_2$	2015	0.18127	0.17994	[ 0.17236 , 0.19516 ]
$\overline{B}_2$	2015	0.06680	0.06685	[ 0.06333 , 0.07060 ]

Table 7: The 95% confidence interval prediction corresponding to the First and Second Stage of *Bachillerato*, in both, state and private high schools all over Spain during academic years 2009–2010 to 2014–2015. Each row shows the percentage of girls/boys who promote ( $G_i/B_i$ ) and do not promote ( $\overline{G}_i/\overline{B}_i$ ) for each level  $i = 1, 2$ .

Academic year	Total Percentage of abandon
1999–2000	1.58
2000–2001	1.51
2001–2002	1.46
2002–2003	1.41
2003–2004	1.37
2004–2005	1.33
2005–2006	1.30
2006–2007	1.27
2007–2008	1.25
2008–2009	1.22
2009–2010	1.21
2010–2011	1.19
2011–2012	1.17
2012–2013	1.16
2013–2014	1.15
2014–2015	1.13

Table 8: Estimation of the percentage of abandon in Spanish *Bachillerato* during the academic years from 1999–2000 to 2014–2015.

Academic Year	2009–2010	2010–2011	2011–2012	2012–2013	2013–2014	2014–2015
Mean	1.25	1.23	1.21	1.19	1.18	1.16
Median	1.25	1.23	1.21	1.19	1.18	1.16
Percentile 2.5	1.20	1.17	1.15	1.13	1.12	1.10
Percentile 97.5	1.30	1.27	1.25	1.23	1.23	1.21

Table 9: Descriptive analysis of the percentage of abandon in Spanish *Bachillerato* during the academic years from 2009–2010 to 2014–2015.

Group	Time (t)	Real data	Confidence Interval Predicted
$G_1$	2010	0.20923	[ 0.17993 , 0.21227 ]
$\overline{G}_1$	2010	0.07005	[ 0.06512 , 0.07041 ]
$B_1$	2010	0.17037	[ 0.16987 , 0.19554 ]
$\overline{B}_1$	2010	0.07485	[ 0.07564 , 0.08020 ]
$G_2$	2010	0.19525	[ 0.15286 , 0.18575 ]
$\overline{G}_2$	2010	0.06399	[ 0.06574 , 0.07340 ]
$B_2$	2010	0.15286	[ 0.15632 , 0.17254 ]
$\overline{B}_2$	2010	0.06340	[ 0.06852 , 0.07391 ]

Table 10: The 95% confidence interval prediction and the real academic results corresponding to the First and Second Stage of *Bachillerato*, in both, state and private high schools all over Spain during the academic year 2009–2010. Each row shows the rate of girls/boys who promote ( $G_i/B_i$ ) and do not promote ( $\overline{G}_i/\overline{B}_i$ ) for each level  $i = 1, 2$ .