The Determinants of University Students Success: a Bivariate Latent Variable Model

Claudia Pigini

Scientific committee:

Renato Balducci
Marco Gallegati
Alberto Niccoli
Alberto Zazzaro

Collana curata da:
Massimo Tamberi

Abstract

The analysis of performance indicators of university students has become of wide interest expecially in Italy where, over the last few decades, graduation rates have been well below the average of both European and OECD countries. This paper proposes an alternative method to jointly estimate the determinants of students academic success, in terms of both potential credits and retention, one year after they first enrolled and a further analysis to evaluate whether there are any factors significantly determining the probability of dropping out, once we consider the students potential academic performance *ceteris paribus*. We implement the algorithm to estimate the parameters of a bivariate latent variable system and then of a conditional mean equation.

Address: Department of Economics - Università Politecnica delle Marche - c.pigini@univpm.it

The Determinants of University Students Success: a Bivariate Latent Variable Model

Claudia Pigini

1 Introduction and motivation

The analysis of performance indicators of university students has become of wide interest expecially in Italy where, over the last few decades, completion rates have been well below the average of both European and OECD countries [1]. These indicators have mostly been identified with the drop-out rate and the progression rate, being the latter defined as the ratio bewteen credits actually acquired by the student during the first year and the credits the student was supposed to acquire at that same time. The availability of administrative data for some Italian universities lead to the application of several empirical strategies, all aiming to find the determinants of such indicators and whether the university reform, introduced in 2001[2], has improved them. The determinants of student performance, expecially of the drop-out rate, have been classified in six categories by the empirical literature (Di Pietro, Cutillo (2008)): personal, school related, ability related, family related, geographical and attendance related.

Although the use of probability models and treatment evaluation procedures in the existing literature has given consistent results about the determinants of the drop-out and progression rate, we did not find any attempt to evaluate the student's probability of success in terms of both academic progression and attainment. This paper proposes an alternative method to jointly estimate the determinants of students academic success, in terms of both potential credits and retention, one year after they first enrolled and a further analysis to evaluate whether there are any significant factors determining the probability of dropping out, once we consider students "ability" *ceteris paribus*. Academic success is identified with the student's propension not to drop-out and the potential credits he/she may acquire during the first year. Students who drop out are those who do not enroll into the second year. Because Italian universities allow for registration into the second year even if the student has not passed all the exams scheduled for the first one, we are able to observe if he/she continues the studies via enrolment into the second year. We consider a system of two latent variables where the first equation has as the dependent variable the potential credits the students may acquire during the first year, while the second models

---

[1] OECD (2009):Education at a Glance

[2] The ministerial decree n. 509/1999 introduced a new framework regarding the Italian higher education system that came into effect in the academic year 2001/2002. The "Bologna Process" is a result of a series of conferences (Paris 1998, Bologna 1999, Prague 2001, Berlin 2003 and Bergen 2005) whose goal was to develop an integrated and coherent European Higher Education Area (EHEA).

the propensity to enroll into the second year (student's retention). We use the concept of potential credits and we treat it as a latent variable as we do not observe how many credits a student may acquire during the first year for several reasons. First of all we only observe the fixed amount of credits that is assigned to each exam the student has passed. Secondly, the Italian university grading system has a peculiarity that makes it reasonable to choose a latent variable approach for credits: unlike other students, the Italian ones can choose to refuse the grade of an exam they sit and try it again to get a higher one. So actual credits would not be a valid indicator of students success in terms of progression as they do not tell us whether the student would have been able to get more credits if he/she accepted lower grades, if the student failed and if he/she did not sit the exam at all. For these reasons we model the potential credits as a latent variable and for each student we observe an interval whose lower bound is the actual amount of credits while the upper bound is the amount of credits the student would account for if he/she passed one more exam or did not refuse one. The two equations system is estimated by maximum likelihood under the assumption of joint normality of the two equations' error terms. Starting values for maximum likelihood are derived by a two step procedure that first estimates the probability of dropping out and gives the generalized residuals to be included among the covariates of the interval regression at the second stage in order to control for the correlation across the two equations.

Although modelling two separate equations (Boero, Laureti, Naylor (2005)) leads to consistent estimates of coefficients, potential correlation within the two performance indicators is not taken into account. It is in fact plausible that a student's drop out decision might be influenced by his own success in passing exams and therefore acquiring credits. Even though data do not allow for an identification of such causality (we do not know if the drop-out decision is made before or after having passed any exams), the maximum likelihood method includes the estimation of such correlation coefficient and exploits all the information available on the number of credits acquired by the students. Furthermore, a hypotetical policy maker would also find it useful to have information about characteristics of incoming students who, given the same potential performance in terms of passed exams, still have a significant probability of dropping out. That is, once we control for the student's accademic potential, we may find out what influences the drop out decision beyond the six major categories and therefore identify factors that can be acted on to reduce the drop-out rate. To do so, we calculate the paramenters of the mean of the probability of not dropping out conditional on the student potential performance as function of the previous estimation. Using administrative data from four faculties of Università Politecnica delle Marche, first we run a joint estimation of the determinants of both the probability of retention and of getting a certain number of credits by maximum likelihood and secondly we try to find the parameters of the retention probability mean conditional to the student's performance in terms of potential credits as a function of the parameters of the previous estimation [3].

The paper is organized as follows: section 2 reviews both traditional and more recent works about the evaluation of students' performance, section 3 outlines the methodology, section 4 describes the data set and section 5 reports

---

[3]We implement the algorithm to estimate the parameters of the bivariate latent variable system and then of the conditional mean equation in Gretl using version 1.8.7.

the estimation results, section 6 concludes.

2 Literature review

To the best of our knowledge, literature on the Italian case focuses on university students performance and drop-out mainly to evaluate if the "3+2" reform, that followed the Bologna declaration, has positively affected them. The case of Italy has been widely discussed as the reform, effective since 2001, had the reduction of both the drop-out rate and graduation time among its primary goals. Common findings are that individual characteristics such as gender, age and prior school background are significant in determining the probability of dropping out and the student performance in terms of passed exams.

Boero, Laureti, Naylor (2005) run two separate estimations for the drop-out probability and the progression rate. They estimate the probability of dropping out with a Probit model and the progression rate by a regression on its logistic transformation as it is bounded between zero and one. Results line up with the expected ones but do not take into account correlation between the two indicators.

Bivariate models have been used in the more traditional literature on drop-out but only to correct for a selection bias that generates when not considering the decision of enrolling after leaving high school. Di Pietro (2004) implements a bivariate probability model with sample selection to account for the two sequential decisions of enrolling and dropping out and he included family background and labour market condition among the covariates that turn out to be significant. There are similar findings in Cingano, Cipollone (2007) who use a type II Tobit model on survey data to account for both the decision of enrolling and dropping out afterwards. They find out that family and educational background are determinant for both phenomena.

Several procedures have been applyed to control for students characteristics and isolate the effect of the university reform on drop-out and student performance. Bratti, Broccolini, Staffolani (2006) use propensity score matching to evaluate the impact of the reform on students performance as well as D'Hombres (2007) who studies such effects on the drop-out and on students status (active vs inactive) instead. Di Pietro and Cutillo (2008) use the Oaxaca decomposition method on three students cohorts to study if the university reform has changed the average predicted conditional probability of dropping out.

Finally Belloc, Maruotti and Petrella (2009) analyse the drop out rate and its determinants performing a generalized linear mixed model estimation and including, among the covariates, students performance (divided in four classes) measured by an index based on the exam grades and the number of credits associated with each exam.

Starting from the existing literature, the methodology we propose allows for the joint estimation of the drop-out probability and student performance accounting for the correlation across them. Unlike the work of Belloc, Maruotti and Petrella, our model does not include average exams grade as university performance because of the problems that would arise if we considered that the grade we observe is sensitive to endogenous truncation due to the possibility of refusing grades, of being rejected (if a student does not pass an exam we do not observe the grade he/she has been rejected with) or of not sitting the exam at

all. Furthermore, it offers the possibility to evaluate what really influences the drop-out decision, once we control for those characteristics that influence both drop-out and performance.

3 Methodology

3.1 Model specification

We analyse a system of two latent variables: the first one, x_i^*, is the potential credits a student can acquire during the first year and the second one, y_i^*, is the propensity to enroll into the second year.

$$x_i^* = z_i'\beta + \varepsilon_i \tag{1}$$

$$y_i^* = z_i'\gamma + v_i \tag{2}$$

Equation (1) will be referred to as the potential credits equation and (2) as the retention equation, where z_i is a set of covariates including individual characteristics and prior school performance (sex, age, high school grade and provenance, region of residence, enrolment year). Key assumption of the model is the joint normality of the error terms ε_i and v_i:

$$\begin{pmatrix} \varepsilon_i \\ v_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \sigma_\varepsilon^2 & \rho\sigma_\varepsilon \\ \rho\sigma_\varepsilon & 1 \end{pmatrix} \right)$$

Let x_i^* be normally distributed as the marginal density of error term:

$$\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$$

and x_i be observed only as the lower bound, a_j, of one of the J intervals x_i^* belongs to. The intervals lower bounds are the credits the student has acquired while the upper bounds, a_{j+1} are the credits the student would have acquired if he passed one more exam.

$$x_i = a_1 \quad \text{if} \quad x_i^* \in (-\infty, a_2)$$

$$x_i = a_2 \quad \text{if} \quad x_i^* \in [a_2, a_3)$$

$$\vdots$$

$$x_i = a_j \quad \text{if} \quad x_i^* \in [a_j, a_{j+1})$$

$$\vdots$$

$$x_i = a_{J-1} \quad \text{if} \quad x_i^* \in [a_{J-1}, a_J)$$

$$x_i = a_J \quad \text{if} \quad x_i^* \in [a_J, +\infty)$$

The propensity not to drop out y_i^* has also a normal distribution given by:

$$v_i \sim N(0, 1)$$

and we observe a dichotomous variable y_i according to the following rule:

$$y_i = 1 \quad if \quad y_i^* > 0$$

$$y_i = 0 \quad if \quad y_i^* \leq 0$$

that is y_i is equal to 1 if the student stays and equal to 0 if he drops out. So the joint normal distribution of the two latent variables is:

$$\begin{pmatrix} x_i^* \\ y_i^* \end{pmatrix} \sim N\left(\begin{pmatrix} z_i'\beta \\ w_i'\gamma \end{pmatrix} ; \begin{pmatrix} \sigma_\varepsilon^2 & \rho\sigma_\varepsilon \\ \rho\sigma_\varepsilon & 1 \end{pmatrix} \right) \tag{3}$$

We want to estimate jointly the parameters of both the potential credits and the retention equations, along with σ_ε and the correlation coefficient ρ :

$$\theta = \begin{pmatrix} \beta \\ \gamma \\ \sigma_\varepsilon \\ \rho \end{pmatrix}$$

We simultaneously estimate all the parameters with maximum likelihood that uses as starting values coefficients estimated by a two-step procedure.

3.2 Two-step

The two-step estimation is carried out as in Rivers and Vuong (1988). We use equation (2) as a "selection equation". Therefore we consider the distribution of the error term ε_i conditional to v_i:

$$\varepsilon_i | v_i \sim N(\rho\sigma_\varepsilon v_i; \quad \sigma_\varepsilon^2(1-\rho^2))$$

so that we can write ε_i as its best linear predictor plus an error term that follows a standard normal distribution

$$\varepsilon_i = \rho\sigma_\varepsilon v_i + \eta_i$$

$$\eta_i \sim N(0;1)$$

that we can substitute in equation (1) obtaining:

$$x_i^* = z_i'\beta + \rho\sigma_\varepsilon v_i + \eta_i$$

Thus we run the two-step procedure in the following manner: we first estimate a Probit model for the retention equation, (2), and we compute its generalized residuals $\hat{v}_i$ (inverse Mill's ratio) and secondly we substitute them in the potential credits equation, (1), whose parameters can now be estimated with an interval regression:

$$x_i = z_i'\beta + \rho\sigma_\varepsilon \hat{v}_i + \eta_i$$

Finally $\hat{\rho}$ can be derived as the ratio between the coefficient of the predicted generalized residuals and $\hat{\sigma_\varepsilon}$.

The two-step estimation solely is much simpler to implement. However, the coefficients covariance matrix estimator is not consistent unless adjusted as shown in Wooldridge (2002) and this could, therefore, result in misleading diagnostics. The covariance matrix estimator is consistent only when the two error terms are independent, being the two-step procedure simply the estimation of two separate equations. Furthermore, a valid test for independece is verifying that the t-ratio for $\widehat{\rho\sigma_\epsilon}$ in the second step is $|t - \text{ratio}| < 2$.

3.3 Maximum likelihood

Once we ran the two-step procedure, we use its estimated coefficients as starting values to implement a maximum likelihood estimation whose likelihood function is built as follows. Let us first introduce the notation:

$$X_j = \frac{a_j - z_i'\beta}{\sigma_\varepsilon} \qquad X_{j+1} = \frac{a_{j+1} - z_i'\beta}{\sigma_\varepsilon} \qquad Y = -w_i'\gamma$$

where X_j and X_{j+1} are the standarized values of, respectively, the lower and the upper bound of interval j and Y is the standardized value of y_i when equal to zero. We first write the credits interval marginal probability as:

$$Pr(X_j) = \begin{cases} Pr(-\infty < x_i^* < a_2) = A_1 = \Phi(X_2) \quad j = 1 \\ \\ Pr(a_j \leq x_i^* < a_{j+1}) = A_2 = \Phi(X_{j+1}) - \Phi(X_j) \quad j = 2, ..., J - 1 \\ \\ Pr(a_J \leq x_i^* < +\infty) = A_3 = 1 - \Phi(X_J) \quad j = J \end{cases} \quad (4)$$

We can also write the joint probability of being in a certain credit interval and dropping out at the same time, $Pr(X_j, y_i^* \leq 0)$, as:

$$Pr(X_j, y_i^* \leq 0) = \begin{cases} Pr(-\infty < x_i^* < a_2, y_i^* \leq 0) = B_1 = \Phi_2(X_2, Y, \rho) \quad j = 1 \\ \\ Pr(a_j \leq x_i^* < a_{j+1}, y_i^* \leq 0) = B_2 = \Phi_2(X_{j+1}, Y, \rho) + \\ \qquad\qquad -\Phi_2(X_j, Y, \rho) \quad j = 2, ..., J - 1 \\ \\ Pr(a_J \leq x_i^* < +\infty, y_i^* \leq 0) = B_3 = \Phi(Y) - \Phi_2(X_J, Y, \rho) \quad j = J \end{cases} \quad (5)$$

while the joint probability of being in a certain interval and not dropping out, $Pr(X_j, y_i^* > 0)$, can be calculated as the difference beteween (4) and (5):

$$Pr(X_j, y_i^* > 0) = \begin{cases} Pr(-\infty < x_i^* < a_2, y_i^* > 0) = A_1 - B_1 \quad j = 1 \\ \\ Pr(a_j \leq x_i^* < a_{j+1}, y_i^* > 0) = A_2 - B_2 \quad j = 2, ..., J - 1 \\ \\ Pr(a_J \leq x_i^* < +\infty, y_i^* > 0) = A_3 - B_3 \quad j = J \end{cases} \quad (6)$$

where Φ and Φ_2 are respectively the univariate and bivariate standard normal cumulative distribution functions.

The likelihood function can be written as:

$$L(\theta; x_1, ..., x_n) = \prod_{i=1}^{n} Pr(X_j, y_i^* > 0)^{y_i} Pr(X_j, y_i^* \leq 0)^{(1-y_i)} \tag{7}$$

so the log-likelihood function for observation i, the one we are going to maximize, is:

$$\ell_i(\theta; x_1, ..., x_n) = y_i \ln[Pr(X_j, y_i^* > 0)] + (1 - y_i) \ln[Pr(X_j, y_i^* \leq 0)] \tag{8}$$

Assuming that all these observations are independent, we can sum across observations to get the log-likelihood for the complete sample:

$$\ell(\theta; x_1, ..., x_n) = \sum_{i=1}^{n} y_i \ln[Pr(X_j, y_i^* > 0)] + (1 - y_i) \ln[Pr(X_j, y_i^* \leq 0)] \tag{9}$$

3.4   Conditional mean equation

Other than jointly estimating all the parameters of the two equations, we can be interested in finding whether there is still a significant probability of dropping out once we control for the potential credits. While the methodology proposed above already allows for the estimation of the correlation coefficient ρ between the potential credits and the propensity not to drop out, this result does not explain if the decision of dropping out is influenced by some factors itself other than a poor academic performance in terms of credits acquired. A futher analysis might give useful information about the drop-out decision as we do not know if it occurs before or after having passed any exams, expecially if there are no significant differences between the determinants of the two equations. Using the results of the maximum likelihood estimation and assuming that the two sets of covariates are the same across equations, we combine (1) and (2) to obtain the mean of the propensity not to drop out conditional on the student's potential credits, which we do not observe, and the set of covariates. Given the bivariate normal distribution of the two latent variables (3), the univariate normal distribution of y_i^* conditional on x_i^* and the covariates z_i can be written as

$$y_i^* | x_i^* \sim N(z_i'\gamma + \frac{\rho}{\sigma_\varepsilon}(x_i^* - z_i'\beta); 1 - \rho^2)$$

so that the conditional mean is

$$E[y_i^* | x_i^*, z_i] = z_i'\gamma + \frac{\rho}{\sigma_\varepsilon} x_i^* - \frac{\rho}{\sigma_\varepsilon} z_i'\beta$$

$$E[y_i^* | x_i^*, z_i] = \alpha_0 x_i^* + z_i'\alpha_1 \tag{10}$$

where

$$\alpha_0 = \frac{\rho}{\sigma_\varepsilon} \qquad \alpha_1 = \gamma - \frac{\rho}{\sigma_\varepsilon}\beta \tag{11}$$

are the coefficient of the potential credits and the vector of coefficients associated with the other covariates respectively. The idea is that when the results confirm the more traditional empirical evidence, according to which academic

success diminishes for older students and for poorer school backgrounds, coefficients associated with the covariates are very similar across the two equations. However, with a high value of ρ, they may offset each other when we consider the determinants of the probability of dropping out conditional on the academic potential, leading to values of the elements of α to be close to zero. The parameters of equation (10) might be of some interest when, for example, evaluating entry tests of incoming students. As the correlation coefficient among the two equations tells us nothing about causality, we are in fact able, through this equation, to identify the determinants of a student's propensity not to drop out once we controlled for the potential credits he may achieve. This would tell us that there are factors influencing the drop-out decision independently from the performance in terms of credits, and that a hypotetical policy-maker should take into account regardless of the entry test results, provided these results are good predictors of students potential credits.

4 Data set description and algorithm implementation

We use administrative data on enrolments from 1999 to 2006 from the faculties of Agriculture, Economics, Engineering and Sciences of "Universitá Politecnica delle Marche". We consider enrolments into the first year of the first lever degrees (three year) after 2001 and enrolments into the old type of graduation courses (four or five years) for years 1999 and 2000. Table (1) reports the descriptions of the variables used in the estimation.

We use as covariates (z) individual and background characteristics such as gender (1 if male), age at the time of enrolment and its square transformation, high school provenance (we take sciences as benchmark, and then we have classical, languages, technical, pedagogical, vocational and other schools) and the deviation of high school grade from the average grade in the same type of high school attended. We also control for the average peer effect (the average high school grade of students who are attending the same course) and whether the students is enrolled into a peripheral faculty. We include students' geographical provenance (our benchmark is students who live in the same city as the university, then we have the same province, same region, other Italian regions, and other countries) and enrolment year. Descriptive statistics are reported in the appendix.

As already outlined, the two latent variables are the potential credits and the propensity not to drop out of university, so that the system is:

$$credits_i^* = z_i'\beta + \varepsilon_i \qquad (12)$$

$$retent_i^* = w_i'\gamma + \upsilon_i \qquad (13)$$

Being the number of potential credits a latent variable, we consider the credits acquired by each student as grouped data, with the lower bound ($credits1$) as the number of credits the student has acquired during the first year, and the upper bound ($credits2$) as the lower bound plus a minimun amount of ects credits the student would have acquired if he passed one more exam or accepted one more grade. For students enrolled in 1999 and 2000 credits have been imputed for the

Table 1: Variables Description

Variables	Description
Dependent Variables	
ects1	lower bound: number of credits acquired during the first year
ects2	upper bound: lower bound plus a minimun amount of credits (different by courses)
retent	=1 if the student enrolls into the second year
Exogenous Variables	
Gender	=1 if male
Age	age at the time of enrolment/10
Age2	square of: age at the time of enrolment/10
High School Grade	from 60 to 100, according to the Italian high school grading system
H.S.Grade: mean dev.	high school grade mean deviation (by school type)
Av. Peer Effect	average high school grade of students who are attending the same course
Sciences	=1 if Scientific Lyceum, benchmark
Classical	=1 if Classical Lyceum
Languages	=1 if Languages Lyceum
Technical	=1 if Technical Institutes
Pedagogical	=1 if Pedagogical Insitute
Vocational	=1 if Professional Institutes
Other	=1 if Other Intitutes
Site: peripheral	=1 if the faculty is not the city of Ancona
Ancona	=1 if the student lives in the city of Ancona, benchmark
Prov. Ancona	=1 if the student lives in the province of Ancona
Marche	=1 if the student lives in other provinces of Marche
Other Regions	=1 if the student lives in regions other than Marche
Foreign	=1 if the student lives abroad
En. Year 1999	=1 if the student enrolled in 1999, benchmark
En. Year 2000	=1 if the student enrolled in 2000
En. Year 2001	=1 if the student enrolled in 2001
En. Year 2002	=1 if the student enrolled in 2002
En. Year 2003	=1 if the student enrolled in 2003
En. Year 2004	=1 if the student enrolled in 2004
En. Year 2005	=1 if the student enrolled in 2005
En. Year 2006	=1 if the student enrolled in 2006
N. of observations	15,476

same passed exams. As for the retention equation we observe a dummy variable which takes the value 1 if the student enrolls into the second year.

To estimate the parameters of the conditional mean equation, we have to assume the same set of covariates, (let's say z_i) in both equations (12) and (13) so we can write it as:

$$mean_i = \alpha_0 credits_i^* + z_i' \alpha_1 \tag{14}$$

where expressions of α_0 and α_1 were already shown in equation (11). Values of the parameters in equations (12),(13) and (14) that we want to estimate can all be expressed as funcions of:

$$\delta = \begin{cases} \beta \\ \gamma \\ \lambda \\ a \end{cases}$$

where λ and a are parameters transformations discussed in the appendix.

5 Results

Estimation results are reported in the appendix separately for the three equations (12), (13) and (14), and for the estimates $\hat{\sigma}_\varepsilon$ and $\hat{\rho}$. Results are presented for each of the four faculties (Agriculture, Economics, Engineering, Sciences).

At a first glance estimates do no seem to be strongly different neither between equations (12) and (13), nor between faculties. Results also comply with most empirical findings in the existing literature. We observe that, only for the faculty of Economics, gender turns out to have a significant impact in both equations as male students are less productive and less likely to continue their studies than female ones. A decreasing negative effect of age appears in both equations and all faculties: the older the students are when they enrol into the first year, the poorer are their academic performance and their probability of retention. However square age allows for such effect to diminish when students enroll after many years they left high school, possibly due to different reasons that brought the latter to the enrolment choice. As for high school background, final grades have a positive impact on the probability of success for students of all faculties while the average peer effect turns out to have negative one for the faculties of Agriculture and Sciences. The impacts of high school types attended almost do not differ within equations for each faculty. For the faculty of Agriculture (Table 6 and Table 7) students who come from a technical, a vocational or form the *Other* type of institutes have a lower probability of succeding in the first year than those who come from a scientific lyceum. In the faculty of Economics and Engineering all students perform worse than those who came from a scientific lyceum except those who came from a classical one that in the faculty of Economics behave the same (Tables 10, 11, 14, and 15). As for the faculty of Sciences (Tables 18 and 19), students form scientific and languages schools have a higher probability of success than the other students. Enrolling into a peripheral site has a positive effect on students' success for Economics and Engineering but not for Agriculture. The impact of geographical

provenance is quite different across faculties. In the faculty of Agriculture, students perform worse than students who live in the city of Ancona only if they live in its province. In the faculty of Economics, students who come from other provinces of region Marche and from other countries have a higher probability of retention while students who come from other regions get less credits, all with respect to those who live in Ancona. While provenance does not make any difference in the probability of retention, for students of the faculty of Engineering living in the province of Ancona and in the other provinces or the region of Marche has respectively a lower and a higher impact on progression than living in Ancona. Students of the faculty of Sciences have all a higher probability of succeding if they do not live in the city of Ancona. The reform introduced in 2001 seems to have boosted students performance in terms of potential credits with respect to year 1999 producing a constant effect over the years. It also has constantly increased the probability of retention for the faculties of Economics and Engineering and only until 2004 for the faculty of Sciences. It appears that the reform had no effect on the probability of retention for the faculty of Agriculture, other than lower it in 2006 with respect to year 1999.

We also report the estimation results for equation (14) for the four faculties (Tables 9, 13, 17 and 21). The coefficient α_0 is the effect of the unobserved potential credits, which we may consider a proxy for student's ability, on the probability of retention. Through these results we want to investigate whether characteristics such as personal, high school related and geographical still have a significant effect on the probability of retention once we control for student's ability. Considering the same amount of potential credits, students' gender and age no longer affect the probability of retention except for the faculty of Engineering. It is, on the contrary, particularly interesting the negative signs of coefficients associated with high school grades, that is, relatively to the prior school experience, students who performed better in high school, when facing poor academic results, are more likely to drop out of university. Only students who attendend a technical or vocational high school in the faculty of Economics and students who come from other insitutes in the faculty of Sciences have still a lower probability of retention than those who come from a scientific lyceum. Being enrolled only into a peripheral site of the faculty of Engineering makes it less likely to drop out and geographical provenance still influences the probability of retention only for the faculty of Economics: considering students' ability *ceteris paribus*, not living in the city of Ancona has a positive effect on the probability of continuing university studies. Furthermore, we find out that given the same potential credits, students who enrolled before the university reform were less likely to drop out than those who did after. This phenomenon could be due to a deterioration of high school background of enrolled students brought by the reform (Bratti, Broccolini, Staffolani (2006)) as a consenquence of the shortening of courses' legal lenght. This innovation made it possible for high school leavers, who would not have countinued their studies for other four or five years, to get a first level degree after only three years of university studies. This lead to a higher number of enrolled students with a weaker prior school performance that could have boosted the drop-out afterwards. It is also not unrealistic to think that the reform has made it less difficult for students to modify their university choice after the first year as credits already acquired are transferable to the new academic carrier.

11

6 Conclusions

The paper has developed an alternative empirical methodology to estimate the determinants of university students' success considering their first year performance, in terms of both potential credits thay may acquire passing exams during the first year and probability of retention. Using administrative data on four faculties of "Università Politecnica delle Marche", we estimated the parameters of a bivariate latent variable system of two equations by maximum likelihood that, unlike a simpler two-step procedure, gives consistent result in presence of a strong correlation between the two variables. A further transformations of the estimated parameters allows for an analysis where we attempt to find determinants of the drop-out probability once we controlled for the students' potential credits which we may consider as a proxy for students' ability. While the results obtained with the estimation of the bivariate latent variable system comply with most empirical findings in the existing literature, we find out that potential credits are alone a quite good predictor of the probability of retention that is now only sensitive to the prior school performance and to the introduction of the reform in 2001. With ability *ceteris paribus* a higher performance in high school makes it more likely to drop out, possibly to change one's own study path, and it is reasonable to think that the effect of the deterioration of high school backgrounds due to a higher number of enrolments after 2001, has neutralized the positive effect of the reform itself on the probability of retention. Hypotetically, this last equation would also reveal some useful information to a policy maker who would have to choose which students to enroll. If an entry test said something about the academic potential success of an incoming student and its result were likely to be in line with the student's prior school performance, our last findings would correct such result that could, in some cases, be misleading.

References

Belloc F., Maruotti A., Petrella L., (2009) "University drop-out: an Italian experience", *Higher Education*, vol. 60, no. 2, 127-138

Bhat C. R., (1994), "Imputing a continuous income variable from grouped and missing income observations", *Economic Letters* 46 (1994) 311-319

Boero G., Laureti T., Naylor R. (2005), " An econometric analysis of student withdrawal and progression in post reform Italian universities", CRENOS Working Papers, no. 1048

Bratti M., Broccolini C., Staffolani S. (2006), " Is "3+2" equal to four? On the effects of university reform in Italy", Università Politecnica delle Marche, Dipartimento di economia Working Papers 251

Bratti M., Broccolini C., Staffolani S. (2010), "Is "3+2" equal to four? Higher education reform, student's time allocation and academic performance in Italy: Evidence from a Faculty of Economics ", *Rivista italiana degli economisti*, forthcoming

Cappellari, L., Lucifora, C. (2008), "The Bologna Process and the College Enrolment Decisions", IZA Discussion Paper No. 3444, IZA, Bonn.

Cipollone, P. and Cingano, F. (2007), "University Drop-out - The case of Italy", Banca d'Italia Temi di discussione (Working Paper) No. 626

D'Hombres, B.(2007), "The impact of university reforms on dropout rates and students' status: Evidence from Italy", JRC Scentific and Technical Reports, European Commission

Davidson, R. and MacKinnon, J. G. (2004), *Econometric Theory and Methods*, New York: Oxford University Press.

Di Pietro, G.(2004), "The determinants of university dropout in Italy: A bivariate probability model with sample selection", *Applied Economics Letters*, 11(3), 187-191

Di Pietro, G., Cutillo, A. (2008), "Degree flexibility and university dropout: The Italian Experience", *Economics of Education review*, 27(5), 546-555

Heckman J.J. (1976), "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for such Models", *Annals of Economics and Social Measurement* 5, 475-492

Heckman J.J. (1979), "Sample Selection Bias as a Specification Error", *Econometrica* 47, 153-161

OECD (2009), "Education at a glance", OECD indicators.

Rivers D., Q.H. Vuong (1988), "Limited information estimators and exogeneity tests for simultaneous probit models", *Journal of Econometrics* 39, 347-366

Rosenbaum, S. (1961), "Moments of a truncated bivariate normal distribution" *Journal of the Royal Statistic Society, Series B, 23, 223-229*

Stern, S. (1991), "Imputing a continuous income variable from a bracketed income variable with special attention to missing observations", *Economic Letters* 37, 287-291

Stewart, M.B., (1983), "On least squares estimation when the dependent variable is grouped", *Review of Economic Studies*, 737-753

Wooldridge J. M. (2002), *Econometric Analysis of Cross Section and Panel Data*, Cambridge, Massachussets, MIT Press

Appendix

Descriptive Statistics

Table 2: Variables Descriptive Statistics (mean) by Faculty

Variables	Agriculture	Economics	Engineering	Sciences
Dependent Variables				
Retention Rate	0.76	0.83	0.74	0.59
Average credits	24.84	31.26	24.97	13.94
Exogenous Variables				
Gender	0.65	0.45	0.83	0.36
Age	20.17	19.83	19.80	19.93
Age2	1.42	1.41	1.40	1.41
High School Grade	78.60	81.92	83.82	79.32
Av. Peer Effect	78.48	81.65	83.68	79.36
Sciences	0.32	0.29	0.39	0.49
Classical	0.05	0.05	0.04	0.13
Languages	0.03	0.04	0.01	0.05
Technical	0.45	0.50	0.50	0.19
Pedagogical	0.03	0.04	0.01	0.07
Vocational	0.12	0.07	0.05	0.06
Other	0.01	0.00	0.01	0.01
Site: peripheral	0.01	0.13	0.13	0.00
Ancona	0.17	0.16	0.08	0.13
Prov. Ancona	0.36	0.35	0.24	0.33
Marche	0.34	0.39	0.40	0.31
Other Regions	0.12	0.08	0.27	0.22
Foreign	0.00	0.02	0.01	0.01
En. Year 1999	0.07	0.08	0.09	0.09
En. Year 2000	0.08	0.09	0.09	0.10
En. Year 2001	0.12	0.13	0.13	0.09
En. Year 2002	0.10	0.14	0.14	0.12
En. Year 2003	0.12	0.14	0.14	0.15
En. Year 2004	0.17	0.15	0.14	0.14
En. Year 2005	0.18	0.14	0.13	0.13
En. Year 2006	0.15	0.13	0.13	0.17
N. of observations	1099.	5246.	7272.	1859.

Table 3: Retention Rate Specifics

Variables	Agriculture	Economics	Engineering	Sciences
Male	0.75	0.79	0.72	0.54
Female	0.77	0.85	0.81	0.62
Age 19-21	0.79	0.85	0.77	0.61
Age 22-30	0.57	0.54	0.33	0.45
Age over 30	0.30	0.53	0.31	0.43
High School Grade 60-70	0.62	0.70	0.54	0.46
High School Grade 70-80	0.78	0.81	0.68	0.58
High School Grade 80-90	0.82	0.88	0.78	0.65
High School Grade 90-100	0.88	0.90	0.85	0.73
Average Peer Effect 60-80	0.77	0.76	0.60	0.62
Average Peer Effect 80-90	0.73	0.84	0.76	0.58
Scientific	0.83	0.89	0.85	0.68
Classical	0.81	0.85	0.76	0.64
Languages	0.83	0.79	0.70	0.62
Technical	0.76	0.81	0.68	0.46
Pedagogical	0.69	0.82	0.62	0.51
Vocational	0.55	0.67	0.44	0.36
Other	0.50	0.80	0.46	0.39
Site: peripheral	0.57	0.77	0.69	-.
Site: Ancona	0.76	0.83	0.75	0.59
Ancona	0.76	0.79	0.74	0.53
Prov Ancona	0.70	0.83	0.73	0.61
Marche	0.82	0.85	0.75	0.56
Other Regions	0.76	0.79	0.72	0.64
Foreign	-.	0.74	0.58	0.63
En Year 1999	0.86	0.87	0.77	0.59
En Year 2000	0.79	0.84	0.76	0.67
En Year 2001	0.69	0.85	0.75	0.63
En Year 2002	0.75	0.83	0.76	0.68
En Year 2003	0.81	0.85	0.76	0.67
En Year 2004	0.74	0.78	0.68	0.61
En Year 2005	0.76	0.82	0.75	0.47
En Year 2006	0.73	0.79	0.70	0.48

Table 4: Credits specifics for students who drop out, by Faculty

Variables	Agriculture	Economics	Engineering	Sciences
Male	3.51	4.93	3.29	2.30
Female	4.88	5.21	4.22	3.49
Age 19-21	4.62	5.97	3.82	3.25
Age 22-30	2.00	1.38	1.91	1.26
Age over 30	0.81	2.06	0.73	0.62
High School Grade 60-70	2.83	3.63	2.08	1.41
High School Grade 70-80	5.52	5.36	2.99	2.40
High School Grade 80-90	3.32	5.34	3.81	3.68
High School Grade 90-100	6.55	7.81	5.74	7.40
Av Peer Effect 60-80	3.87	4.36	2.82	1.83
Av Peer Effect 80-90	4.22	5.34	3.55	3.43
Scientific	6.57	6.28	4.80	3.92
Classical	5.60	8.22	2.49	4.96
Languages	1.20	6.29	6.68	1.97
Technical	3.20	4.92	3.22	1.66
Pedagogical	4.45	3.22	3.93	3.15
Vocational	2.60	3.31	1.56	1.05
Other	8.67	0.00	2.93	3.00
Site: peripheral	0.00	4.54	2.98	1.58
Site: Ancona	4.07	5.17	3.48	-.
Ancona	3.98	5.02	2.68	2.63
Prov Ancona	3.93	5.32	3.41	2.11
Marche	2.46	5.26	3.74	3.41
Other Regions	7.27	3.83	3.10	4.04
Foreign	-.	3.73	4.47	3.43
En Year 1999	0.91	2.35	0.86	1.14
En Year 2000	0.28	1.43	1.25	2.85
En Year 2001	0.71	4.46	2.75	1.30
En Year 2002	6.48	3.51	3.23	0.68
En Year 2003	3.35	5.26	2.51	1.06
En Year 2004	3.06	5.62	4.59	4.38
En Year 2005	5.81	8.05	5.69	4.91
En Year 2006	7.13	6.27	4.09	4.34

Table 5: Credits specifics for students who enroll into the second year, by Faculty

Variables	Agriculture	Economics	Engineering	Sciences
Male	30.87	34.35	31.82	19.88
Female	32.67	38.58	36.29	22.22
Age 19-21	32.16	37.36	33.09	21.69
Age 22-30	24.04	26.59	20.83	21.02
Age over 30	13.44	23.34	13.33	10.00
High School Grade 60-70	23.35	27.39	20.80	14.40
High School Grade 70-80	29.75	33.99	26.92	20.01
High School Grade 80-90	34.44	38.30	31.86	22.74
High School Grade 90-100	39.75	43.46	39.84	27.39
Average Peer Effect 60-80	31.32	34.85	34.86	22.66
Average Peer Effect 80-90	32.05	37.24	32.37	20.94
Scientific	32.53	37.00	35.49	23.47
Classical	30.56	36.27	28.64	18.48
Languages	34.80	36.06	29.87	23.42
Technical	32.20	37.58	30.89	18.68
Pedagogical	31.83	36.54	25.40	18.29
Vocational	23.74	30.33	25.14	16.21
Other	10.33	28.85	24.75	20.43
Site: peripheral	31.88	36.64	36.20	-.
Site: Ancona	31.51	36.79	32.13	21.13
Ancona	32.73	36.23	32.52	19.63
Prov Ancona	30.68	36.09	33.23	21.56
Marche	31.65	38.43	33.73	20.60
Other Regions	31.69	33.26	30.66	23.34
Foreign	-.	31.06	25.95	19.67
En Year 1999	13.97	21.18	13.44	17.43
En Year 2000	8.23	20.11	14.29	25.13
En Year 2001	34.58	37.03	35.78	17.98
En Year 2002	39.84	41.46	38.21	16.97
En Year 2003	33.98	40.29	37.75	16.28
En Year 2004	33.41	40.96	37.19	23.01
En Year 2005	35.07	40.11	37.49	25.94
En Year 2006	38.16	41.43	35.61	29.54

First partial derivatives

Given the additional parameters as transformations of σ_ε and ρ and the probabilities A_s and B_s

$$a = atanh(\rho) \qquad \lambda = ln(\sigma_\varepsilon)$$

$$A_s = Pr(X_j) \quad B_s = Pr(X_j, y_i^* \leq 0) \quad s = 1, 2, 3$$

the first partial derivatives of the log-likelihood function are:

$$\frac{\partial \ell}{\partial \beta} = y_i \frac{1}{A_s - B_s} \frac{\partial (A_s - B_s)}{\partial X} \frac{\partial X}{\partial \beta} + (1 - y_i) \frac{1}{B_s} \frac{\partial (B_s)}{\partial X} \frac{\partial X}{\partial \beta}$$

$$\frac{\partial \ell}{\partial \gamma} = y_i \frac{1}{A_s - B_s} \frac{\partial (A_s - B_s)}{\partial Y} \frac{\partial Y}{\partial \gamma} + (1 - y_i) \frac{1}{B_s} \frac{\partial (B_s)}{\partial Y} \frac{\partial Y}{\partial \gamma}$$

$$\frac{\partial \ell}{\partial \lambda} = y_i \frac{1}{A_s - B_s} \frac{\partial (A_s - B_s)}{\partial \sigma_\varepsilon} \frac{\partial \sigma_\varepsilon}{\partial \lambda} + (1 - y_i) \frac{1}{B_s} \frac{\partial (B_s)}{\partial \sigma_\varepsilon} \frac{\partial \sigma_\varepsilon}{\partial \lambda}$$

$$\frac{\partial \ell}{\partial a} = y_i \frac{1}{A_s - B_s} \frac{\partial (A_s - B_s)}{\partial a} + (1 - y_i) \frac{1}{B_s} \frac{\partial (B_s)}{\partial a}$$

where:

$$\frac{\partial (B_s)}{\partial X} = \begin{cases} \frac{\partial (B_1)}{\partial X} = \varphi(X_2)\Phi(c_a Y - s_a X) \quad j = 1 \\\\ \frac{\partial (B_2)}{\partial X} = \varphi(X_{j+1})\Phi(c_a Y - s_a X_{j+1}) - \varphi(X_j)\Phi(c_a Y - s_a X_j) \quad j = 2, ..., J - 1 \\\\ \frac{\partial (B_3)}{\partial X} = -\varphi(X_J)\Phi(c_a Y - s_a X_J) \quad j = J \end{cases}$$

$$\frac{\partial (A_s - B_s)}{\partial X} = \begin{cases} \frac{\partial (A_1 - B_1)}{\partial X} = \varphi(X_2) - \frac{\partial (B_1)}{\partial X} \quad j = 1 \\\\ \frac{\partial (A_2 - B_2)}{\partial X} = \varphi(X_{j+1}) - \varphi(X_j) - \frac{\partial (B_2)}{\partial X} \quad j = 2, ..., J - 1 \\\\ \frac{\partial (A_3 - B_3)}{\partial X} = -\varphi(X_J) - \frac{\partial (B_3)}{\partial X} \quad j = J \end{cases}$$

$$\frac{\partial (B_s)}{\partial Y} = \begin{cases} \frac{\partial (B_1)}{\partial Y} = \varphi(Y)\Phi(c_a X_2 - s_a Y) \quad j = 1 \\\\ \frac{\partial (B_2)}{\partial Y} = \varphi(Y)\Phi(c_a X_{j+1} - s_a Y) - \varphi(Y)\Phi(c_a X_j - s_a Y) \quad j = 2, ..., J - 1 \\\\ \frac{\partial (B_3)}{\partial Y} = \varphi(Y) - \varphi(Y)\Phi(c_a X_J - s_a Y) \quad j = J \end{cases}$$

$$\frac{\partial(A_s - B_s)}{\partial Y} = \begin{cases} \frac{\partial(A_1 - B_1)}{\partial Y} = -\frac{\partial(B_1)}{\partial Y} & j = 1 \\[2mm] \frac{\partial(A_2 - B_2)}{\partial Y} = -\frac{\partial(B_2)}{\partial Y} & j = 2, ..., J-1 \\[2mm] \frac{\partial(A_3 - B_3)}{\partial Y} = -\frac{\partial(B_3)}{\partial Y} & j = J \end{cases}$$

$$\frac{\partial(B_s)}{\partial \sigma_\varepsilon} = \begin{cases} \frac{\partial(B_1)}{\partial \sigma_\varepsilon} = \frac{\partial(B_1)}{\partial X_2}\frac{\partial X_2}{\partial \sigma_\varepsilon} & j = 1 \\[2mm] \frac{\partial(B_2)}{\partial \sigma_\varepsilon} = \varphi(X_{j+1})\Phi(c_a Y - s_a X_{j+1})\frac{\partial X_{j+1}}{\partial \sigma_\varepsilon} - \varphi(X_j)\Phi(c_a Y - s_a X_j)\frac{\partial X_j}{\partial \sigma_\varepsilon} & j = 2, ..., J-1 \\[2mm] \frac{\partial(B_3)}{\partial \sigma_\varepsilon} = \frac{\partial(B_3)}{\partial X_J}\frac{\partial X_J}{\partial \sigma_\varepsilon} & j = J \end{cases}$$

$$\frac{\partial(A_s - B_s)}{\partial \sigma_\varepsilon} = \begin{cases} \frac{\partial(A_1 - B_1)}{\partial \sigma_\varepsilon} = [\varphi(X_2) - \frac{\partial(B_1)}{\partial X_2}]\frac{\partial X_2}{\partial \sigma_\varepsilon} & j = 1 \\[2mm] \frac{\partial(A_2 - B_2)}{\partial \sigma_\varepsilon} = \varphi(X_{j+1})(1 - \Phi(c_a Y - s_a X_{j+1}))\frac{\partial X_{j+1}}{\partial \sigma_\varepsilon} + \\ \quad\quad + \varphi(X_j)(1 - \Phi(c_a Y - s_a X_j))\frac{\partial X_j}{\partial \sigma_\varepsilon} & j = 2, ..., J-1 \\[2mm] \frac{\partial(A_3 - B_3)}{\partial \sigma_\varepsilon} = [-\varphi(X_J) - \frac{\partial(B_3)}{\partial X_J}]\frac{\partial X_J}{\partial \sigma_\varepsilon} & j = J \end{cases}$$

$$\frac{\partial(B_s)}{\partial a} = \begin{cases} \frac{\partial(B_1)}{\partial a} = \frac{1}{c_a}\varphi_2(X_2, Y, a) & j = 1 \\[2mm] \frac{\partial(B_2)}{\partial a} = \frac{1}{c_a}[\varphi_2(X_{j+1}, Y, a) - \varphi_2(X_j, Y, a)] & j = 2, ..., J-1 \\[2mm] \frac{\partial(B_3)}{\partial a} = -\frac{1}{c_a}\varphi_2(X_j, Y, a) & j = J \end{cases}$$

$$\frac{\partial(A_s - B_s)}{\partial a} = \begin{cases} \frac{\partial(A_1 - B_1)}{\partial a} = -\frac{\partial(B_1)}{\partial a} & j = 1 \\[2mm] \frac{\partial(A_2 - B_2)}{\partial a} = -\frac{\partial(B_2)}{\partial a} & j = 2, ..., J-1 \\[2mm] \frac{\partial(A_3 - B_3)}{\partial a} = -\frac{\partial(B_3)}{\partial a} & j = J \end{cases}$$

$$\frac{\partial X_j}{\partial \beta} = \frac{\partial(\frac{a_j - z_i'\beta}{\sigma_\varepsilon})}{\partial \beta} = (-\frac{1}{\sigma_\varepsilon})z_i'$$

$$\frac{\partial Y}{\partial \gamma} = \frac{\partial(-w_i'\gamma)}{\partial \gamma} = -w_i'$$

$$\frac{\partial X_j}{\partial \sigma_\varepsilon} = \frac{\partial(\frac{a_j - z_i'\beta}{\sigma_\varepsilon})}{\partial \sigma_\varepsilon} = -\frac{a_j - z_i'\beta}{\sigma_\varepsilon^2}$$

$$\frac{\partial \sigma_\varepsilon}{\partial \lambda} = e^\lambda$$

Estimation Procedure

We run the estimation in Gretl 1.8.7 by calling the function `doestimation`, which we wrote to implement the whole procedure. The fuction takes as arguments the dependent variables, $ects1$ $ects2$ and $retent$, and the two lists of covariates, z and w. All the functions we call inside `doestimation` we have previuosly written. The function proceeds by taking the following steps:

- First it checks for missing observations in values of covariates and restricts the sample. It also transforms lists into matrices.

- It calls the function `twostep` that runs the two-step procedure and returns a vector: it first estimates a probit model for equation (13), it computes its generalized residuals and stores them in order to include them among the explanatory variables of the interval regression it runs afterwards for equation (12). It now creates the vector of parameters we will use as starting values for the maximum likelihood estimation putting together the estimates of coefficients associated to the covariates with the standard deviation and the correlation coefficient, derived as the ratio between the generalized residuals estimated coefficient and the standard deviation:

- It then calls the function `checktheta` that corrects values of the standard deviation and the correlation coefficient across iterations so that critical values do not occur.

- The function runs the maximum likelihood estimation using the log-likelihood series created in the function `loglik` that, among the arguments described above, takes also the coefficient vector produced by the `twostep` function to use as starting values for the maximization. The log-likelihood function has been calculated as in equation (8). The estimation is carried out using analytical first derivatives (reported in Appendix) created in the function `fderiv` and called in the maximum likelihood command. For both the log-likelihood and the first derivatives we use transformations of the parameters σ_ε and ρ (although values of $\hat{\sigma}$ and $\hat{\rho}$ are reported):

$$\lambda = \ln(\sigma_\varepsilon) \qquad a = atanh(\rho)$$

- Coefficient standard errors are calculated via deltha method due to the parameters transformations, as the square root of diagonal elements of the matrix $\widehat{Var(\hat{\theta})}$:

$$\widehat{Var(\hat{\theta})} = J\widehat{V(\hat{\delta})}J'$$

where matrix J is the Jacobian matrix for the function:

$$\hat{\theta} = g(\hat{\delta})$$

whose elemets are its first partial derivatives:

$$J_{ij} = \frac{\partial g(\hat{\delta_i})}{\partial \hat{\delta_j}}$$

For the non transformed parameters, first derivatives are zero other then when the parameter is derived with respect to itself and in that case the result is 1. For the transformed parameters we have:

$$\frac{\partial \hat{\sigma}_\varepsilon}{\partial \hat{\lambda}} = e^{\hat{\lambda}}$$

$$\frac{\partial \hat{\rho}}{\partial \hat{a}} = \frac{1}{c_{\hat{a}}^2}$$

The values of standard errors we present in section 6 are base on the quasi-ML "sandwich" estimator (Davidson, MacKinnon 2004), $\widehat{Var_R}(\hat{\delta})$

$$\widehat{Var_R}(\hat{\delta}) = H(\hat{\delta})^{-1}(G'(\hat{\delta})G(\hat{\delta}))^{-1}H(\hat{\delta})^{-1}$$

However, `doestimation` provides other two estimates of the covariance matrix: the first one is based on the Outer Product of the Gradient:

$$\widehat{Var_{OPG}}(\hat{\delta}) = (G'(\hat{\delta})G(\hat{\delta}))^{-1}$$

while the second covariance matrix is computed from a numerical approximation to the Hessian at convergence:

$$\widehat{Var_H}(\hat{\delta}) = (H(\hat{\delta}))^{-1}$$

- A further transformation of parameters is implemented to give coefficients for the conditional mean equation (10) as outlined in section (4.3). The standard errors are based on the robust covariance matrix estimator and Jacobian matrix uses the following first partial derivatives:

$$\frac{\partial \alpha_0}{\partial \beta_j} = 0 \qquad \frac{\partial \alpha_0}{\partial \gamma_j} = 0$$

$$\frac{\partial \alpha_0}{\partial \lambda} = -\alpha_0$$

$$\frac{\partial \alpha_0}{\partial a} = \frac{1}{\sigma_\varepsilon}\frac{1}{c_{\hat{a}}^2}$$

$$\frac{\partial \alpha_{1i}}{\partial \beta_j} = 0 \qquad \frac{\partial \alpha_{1i}}{\partial \beta_i} = -\alpha_0$$

$$\frac{\partial \alpha_{1i}}{\partial \gamma_j} = 0 \qquad \frac{\partial \alpha_{1i}}{\partial \gamma_i} = 1$$

$$\frac{\partial \alpha_1}{\partial \lambda} = \alpha_0\beta$$

$$\frac{\partial \alpha_1}{\partial a} = -\frac{1}{\sigma_\varepsilon}\frac{1}{c_{\hat{a}}^2}\beta$$

- Finally `doestimation` prints the output for equations (12) (13) and (14) using the gretl command `modprint`

Results

Table 6: Potential Credits Equation: Faculty of Agriculture

	Coefficient	Std. Error	z-stat	p-value
Const.	14.7622	3.23454	4.5639	0.0000
Gender: male	0.254227	0.130392	1.9497	0.0512
Age	-2.05405	1.32240	-1.5533	0.1204
Age2	0.0177153	0.241182	0.0735	0.9414
H.S.Grade: mean dev.	0.0803026	0.00476642	16.8476	0.0000
Av. Peer Effect	-0.118818	0.0342397	-3.4702	0.0005
Classical	-0.226150	0.259512	-0.8714	0.3835
Languages	0.118094	0.295181	0.4001	0.6891
Technical	-0.606092	0.145578	-4.1634	0.0000
Pedagogical	-0.508820	0.355253	-1.4323	0.1521
Vocational	-1.84642	0.246053	-7.5041	0.0000
Other	-2.37433	0.398647	-5.9560	0.0000
Site: peripheral	-1.01406	0.720808	-1.4068	0.1595
Prov. Ancona	-0.387507	0.176524	-2.1952	0.0281
Marche	-0.110633	0.186600	-0.5929	0.5533
Other Regions	-0.0626113	0.223862	-0.2797	0.7797
En. Year 2000	-0.950929	0.188081	-5.0560	0.0000
En. Year 2001	1.61638	0.255052	6.3374	0.0000
En. Year 2002	2.06902	0.233042	8.8783	0.0000
En. Year 2003	2.06126	0.213306	9.6634	0.0000
En. Year 2004	1.71052	0.199227	8.5858	0.0000
En. Year 2005	2.07537	0.193477	10.7267	0.0000
En. Year 2006	1.98491	0.202692	9.7928	0.0000

Table 7: Retention Equation: Faculty of Agriculture

	Coefficient	Std. Error	z-stat	p-value
Const	8.90198	2.20476	4.0376	0.0001
Gender: male	0.112668	0.100110	1.1254	0.2604
Age	−1.84949	0.573531	−3.2247	0.0013
Age2	0.202308	0.0921013	2.1966	0.0281
H.S.Grade: mean dev.	0.0295803	0.00376877	7.8488	0.0000
Av. Peer Effect	−0.0631134	0.0252432	−2.5002	0.0124
Classical	−0.0344400	0.235721	−0.1461	0.8838
Languages	0.0276809	0.227194	0.1218	0.9030
Technical	−0.274474	0.112457	−2.4407	0.0147
Pedagogical	−0.287226	0.245161	−1.1716	0.2414
Vocational	−0.664930	0.160527	−4.1422	0.0000
Other	−0.958307	0.409169	−2.3421	0.0192
Site: peripheral	−0.743662	0.413449	−1.7987	0.0721
Prov. Ancona	−0.221727	0.134296	−1.6510	0.0987
Marche	0.180191	0.143485	1.2558	0.2092
Other Regions	−0.0841921	0.177448	−0.4745	0.6352
En. Year 2000	−0.326907	0.209311	−1.5618	0.1183
En. Year 2001	0.0328539	0.208857	0.1573	0.8750
En. Year 2002	−0.143668	0.208364	−0.6895	0.4905
En. Year 2003	0.174805	0.205111	0.8522	0.3941
En. Year 2004	−0.0571107	0.192222	−0.2971	0.7664
En. Year 2005	−0.00943172	0.188965	−0.0499	0.9602
En. Year 2006	−0.338332	0.192621	−1.7565	0.0790

Table 8: Standard Deviation and Correlation Coefficient: Faculty of Agriculture

	Coefficient	Std. Error	z-stat	p-value
σ_ε	1.81096	0.0506616	35.7462	0.0000
ρ	0.850335	0.0167647	50.7218	0.0000

Table 9: Conditional Mean Equation: Faculty of Agriculture

	Coefficient	Std. Error	z-stat	p-value
α_0	0.469550	0.0136659	34.3594	0.0000
Const.	1.97041	1.81201	1.0874	0.2769
Gender: male	−0.00670415	0.0823846	−0.0814	0.9351
Age	−0.885009	0.666254	−1.3283	0.1841
Age2	0.193989	0.123727	1.5679	0.1169
H.S.Grade: mean dev.	−0.00812580	0.00313217	−2.5943	0.0095
Av. Peer Effect	−0.00732239	0.0198641	−0.3686	0.7124
Classical	0.0717486	0.216241	0.3318	0.7400
Languages	−0.0277702	0.169797	−0.1635	0.8701
Technical	0.0101162	0.0944086	0.1072	0.9147
Pedagogical	−0.0483097	0.200996	−0.2404	0.8101
Vocational	0.202057	0.133996	1.5079	0.1316
Other	0.156561	0.323135	0.4845	0.6280
Site: peripheral	−0.267512	0.215475	−1.2415	0.2144
Prov. Ancona	−0.0397737	0.112024	−0.3550	0.7226
Marche	0.232139	0.118968	1.9513	0.0510
Other Regions	−0.0547929	0.151495	−0.3617	0.7176
En. Year 2000	0.119601	0.194248	0.6157	0.5381
En. Year 2001	−0.726116	0.181584	−3.9988	0.0001
En. Year 2002	−1.11518	0.183034	−6.0927	0.0000
En. Year 2003	−0.793057	0.183136	−4.3304	0.0000
En. Year 2004	−0.860285	0.172483	−4.9876	0.0000
En. Year 2005	−0.983922	0.169996	−5.7879	0.0000
En. Year 2006	−1.27035	0.173416	−7.3254	0.0000

Table 10: Potential Credits Equation: Faculty of Economics

	Coefficient	Std. Error	z-stat	p-value
Const	6.45083	2.21861	2.9076	0.0036
Gender:male	−0.208359	0.0660799	−3.1531	0.0016
Age	−7.61913	0.807610	−9.4342	0.0000
Age2	1.04502	0.137416	7.6048	0.0000
H.S.Grade: mean dev.	0.0743434	0.00260541	28.5343	0.0000
Av. Peer Effect	0.0773302	0.0232238	3.3298	0.0009
Classical	−0.0438684	0.149629	−0.2932	0.7694
Languages	−0.750802	0.171713	−4.3724	0.0000
Technical	−0.323158	0.0702658	−4.5991	0.0000
Pedagogical	−0.843756	0.176033	−4.7932	0.0000
Vocational	−1.59943	0.138963	−11.5097	0.0000
Other	−1.35724	0.462687	−2.9334	0.0034
Site: peripheral	0.516664	0.159723	3.2347	0.0012
Prov. Ancona	−0.0396997	0.0938924	−0.4228	0.6724
Marche	0.100030	0.0924730	1.0817	0.2794
Other Regions	−0.448655	0.136385	−3.2896	0.0010
Foreign	0.158432	0.283609	0.5586	0.5764
En. Year 2000	−0.0585324	0.104973	−0.5576	0.5771
En. Year 2001	2.18176	0.105414	20.6970	0.0000
En. Year 2002	2.44489	0.113047	21.6272	0.0000
En. Year 2003	2.45050	0.110079	22.2613	0.0000
En. Year 2004	2.40604	0.115200	20.8857	0.0000
En. Year 2005	2.27692	0.111481	20.4242	0.0000
En. Year 2006	2.34530	0.117495	19.9608	0.0000

Table 11: Retention Equation: Faculty of Economics

	Coefficient	Std. Error	z-stat	p-value
Const	3.15778	1.33046	2.3734	0.0176
Gender: male	−0.121647	0.0458936	−2.6506	0.0080
Age	−2.77304	0.387104	−7.1635	0.0000
Age2	0.388450	0.0646304	6.0103	0.0000
H.S.Grade: mean dev.	0.0234248	0.00183224	12.7848	0.0000
Av. Peer Effect	0.0211494	0.0145689	1.4517	0.1466
Classical	−0.153255	0.110699	−1.3844	0.1662
Languages	−0.441049	0.115451	−3.8202	0.0001
Technical	−0.259199	0.0518558	−4.9985	0.0000
Pedagogical	−0.428003	0.113773	−3.7619	0.0002
Vocational	−0.761614	0.0848841	−8.9724	0.0000
Other	−0.172728	0.251819	−0.6859	0.4928
Site: peripheral	0.189639	0.0984161	1.9269	0.0540
Prov. Ancona	0.0880335	0.0610624	1.4417	0.1494
Marche	0.151139	0.0616434	2.4518	0.0142
Other Regions	−0.0340386	0.0858473	−0.3965	0.6917
Foreign	0.308632	0.176273	1.7509	0.0800
En. Year 2000	0.0536895	0.0876460	0.6126	0.5402
En. Year 2001	0.374260	0.0865238	4.3255	0.0000
En. Year 2002	0.411612	0.0877723	4.6895	0.0000
En. Year 2003	0.442055	0.0885101	4.9944	0.0000
En. Year 2004	0.234476	0.0913882	2.5657	0.0103
En. Year 2005	0.219319	0.0897153	2.4446	0.0145
En. Year 2006	0.158430	0.0893857	1.7724	0.0763

Table 12: Standard Deviation and Correlation Coefficient: Faculty of Economics

	Coefficient	Std. Error	z-stat	p-value
σ_ε	2.10878	0.0308650	68.3226	0.0000
ρ	0.834623	0.00966831	86.3256	0.0000

Table 13: Conditional Mean Equation: Faculty of Economics

	Coefficient	Std. Error	z-stat	p-value
α_0	0.395785	0.00633883	62.4382	0.0000
Const	0.604638	1.03410	0.5847	0.5587
Gender: male	-0.0391812	0.0389286	-1.0065	0.3142
Age	0.242504	0.302855	0.8007	0.4233
Age2	-0.0251521	0.0514561	-0.4888	0.6250
H.S.Grade: mean dev.	-0.00599926	0.00154939	-3.8720	0.0001
Av Peer Effect	-0.00945672	0.0114508	-0.8259	0.4089
Classical	-0.135892	0.100045	-1.3583	0.1744
Languages	-0.143893	0.0944756	-1.5231	0.1277
Technical	-0.131298	0.0453511	-2.8951	0.0038
Pedagogical	-0.0940568	0.0921240	-1.0210	0.3073
Vocational	-0.128584	0.0638357	-2.0143	0.0440
Other	0.364449	0.165980	2.1957	0.0281
Site: peripheral	-0.0148485	0.0790021	-0.1880	0.8509
Prov. Ancona	0.103746	0.0502801	2.0634	0.0391
Marche	0.111549	0.0510291	2.1860	0.0288
Other Regions	0.143533	0.0689846	2.0806	0.0375
Foreign	0.245927	0.132840	1.8513	0.0641
En. Year 2000	0.0768558	0.0785716	0.9782	0.3280
En. Year 2001	-0.489249	0.0757483	-6.4589	0.0000
En. Year 2002	-0.556040	0.0748663	-7.4271	0.0000
En. Year 2003	-0.527818	0.0768315	-6.8698	0.0000
En. Year 2004	-0.717799	0.0794895	-9.0301	0.0000
En. Year 2005	-0.681853	0.0781789	-8.7217	0.0000
En. Year 2006	-0.769805	0.0754718	-10.1999	0.0000

Table 14: Potential Credits Equation: Faculty of Engineering

	Coefficient	Std. Error	z-stat	p-value
Const	10.4821	1.89776	5.5234	0.0000
Gender: male	0.00473948	0.0741672	0.0639	0.9490
Age	−5.52073	0.978836	−5.6401	0.0000
Age2	0.700063	0.175276	3.9941	0.0001
H.S.Grade: mean dev.	0.0887505	0.00225870	39.2928	0.0000
Av. Peer Effect	−0.0117098	0.0142861	−0.8197	0.4124
Classical	−0.964826	0.141232	−6.8315	0.0000
Languages	−1.48828	0.272844	−5.4547	0.0000
Technical	−1.27294	0.0586735	−21.6952	0.0000
Pedagogical	−1.88218	0.342361	−5.4976	0.0000
Vocational	−2.98449	0.157044	−19.0042	0.0000
Other	−2.91013	0.263889	−11.0279	0.0000
Site: peripheral	0.381153	0.112127	3.3993	0.0007
Prov. Ancona	0.0628910	0.107468	0.5852	0.5584
Marche	0.304869	0.103307	2.9511	0.0032
Other Regions	−0.104219	0.106064	−0.9826	0.3258
Foreign	0.291092	0.265283	1.0973	0.2725
En. Year 2000	0.203144	0.0819923	2.4776	0.0132
En. Year 2001	2.46051	0.0958343	25.6746	0.0000
En. Year 2002	2.74917	0.0985760	27.8888	0.0000
En. Year 2003	2.48920	0.0962505	25.8617	0.0000
En. Year 2004	2.13024	0.0947134	22.4915	0.0000
En. Year 2005	2.48397	0.0943838	26.3178	0.0000
En. Year 2006	2.03427	0.0943128	21.5694	0.0000

Table 15: Retention Equation: Faculty of Engineering

	Coefficient	Std. Error	z-stat	p-value
Const	3.26741	1.04715	3.1203	0.0018
Gender: male	0.00159404	0.0475852	0.0335	0.9733
Age	-2.74434	0.522375	-5.2536	0.0000
Age2	0.378182	0.0933305	4.0521	0.0001
H.S.Grade: mean dev.	0.0301530	0.00139729	21.5797	0.0000
Av. Peer Effect	0.0167898	0.00794175	2.1141	0.0345
Classical	-0.338141	0.0880787	-3.8391	0.0001
Languages	-0.653261	0.173970	-3.7550	0.0002
Technical	-0.544614	0.0373783	-14.5703	0.0000
Pedagogical	-0.830394	0.201096	-4.1293	0.0000
Vocational	-1.22964	0.0840289	-14.6336	0.0000
Other	-1.24718	0.143584	-8.6860	0.0000
Site: peripheral	0.245483	0.0628209	3.9077	0.0001
Prov. Ancona	-0.0484796	0.0676357	-0.7168	0.4735
Marche	0.0890438	0.0650259	1.3694	0.1709
Other Regions	-0.0497806	0.0665587	-0.7479	0.4545
Marche	0.0415400	0.161593	0.2571	0.7971
En. Year 2000	0.0269698	0.0599191	0.4501	0.6526
En. Year 2001	0.493705	0.0641206	7.6996	0.0000
En. Year 2002	0.571577	0.0658899	8.6747	0.0000
En. Year 2003	0.486025	0.0660772	7.3554	0.0000
En. Year 2004	0.138320	0.0653087	2.1179	0.0342
En. Year 2005	0.335427	0.0661272	5.0725	0.0000
En. Year 2006	0.170141	0.0648303	2.6244	0.0087

Table 16: Standard Deviation and Correlation Coefficient: Faculty of Engineering

	Coefficient	Std. Error	z-stat	p-value
σ_ε	2.10875	0.0253192	83.2868	0.0000
ρ	0.878268	0.00743363	118.1480	0.0000

Table 17: Conditional Mean Equation: Faculty of Engineering

	Coefficient	Std. Error	z-stat	p-value
α_0	0.416487	0.00556292	74.8684	0.0000
Const	-1.09825	0.655172	-1.6763	0.0937
Gender: male	-0.000379889	0.0382979	-0.0099	0.9921
Age	-0.445032	0.224889	-1.9789	0.0478
Age2	0.0866146	0.0360649	2.4016	0.0163
H.S.Grade: mean dev.	-0.00681038	0.00107886	-6.3125	0.0000
Av. Peer Effect	0.0216668	0.00587033	3.6909	0.0002
Classical	0.0636971	0.0647083	0.9844	0.3249
Languages	-0.0334100	0.139122	-0.2401	0.8102
Technical	-0.0144534	0.0291536	-0.4958	0.6201
Pedagogical	-0.0464908	0.151407	-0.3071	0.7588
Vocational	0.0133580	0.0554206	0.2410	0.8095
Other	-0.0351453	0.0969157	-0.3626	0.7169
Site: peripheral	0.0867377	0.0456080	1.9018	0.0572
Prov. Ancona	-0.0746728	0.0530798	-1.4068	0.1595
Marche	-0.0379301	0.0509783	-0.7440	0.4569
Other Regions	-0.00637474	0.0522322	-0.1220	0.9029
Foreign	-0.0796960	0.115944	-0.6874	0.4919
En. Year 2000	-0.0576370	0.0496577	-1.1607	0.2458
En. Year 2001	-0.531064	0.0490369	-10.8299	0.0000
En. Year 2002	-0.573416	0.0504067	-11.3758	0.0000
En. Year 2003	-0.550694	0.0498958	-11.0369	0.0000
En. Year 2004	-0.748900	0.0513005	-14.5983	0.0000
En. Year 2005	-0.699116	0.0525884	-13.2941	0.0000
En. Year 2006	-0.677107	0.0510252	-13.2700	0.0000

Table 18: Potential Credits Equation: Faculty of Sciences

	Coefficient	Std. Error	z-stat	p-value
Const	21.4269	3.47234	6.1707	0.0000
Gender: male	0.0257308	0.0945835	0.2720	0.7856
Age	-1.53479	0.976325	-1.5720	0.1159
Age2	0.163495	0.160527	1.0185	0.3084
H.S.Grade: mean dev.	0.0593614	0.00342747	17.3193	0.0000
Av. Peer Effect	-0.230290	0.0385432	-5.9748	0.0000
Classical	-0.440669	0.119955	-3.6736	0.0002
Languages	-0.298341	0.194727	-1.5321	0.1255
Technical	-1.23528	0.124453	-9.9257	0.0000
Pedagogical	-0.929865	0.165954	-5.6032	0.0000
Vocational	-1.73861	0.204444	-8.5041	0.0000
Other	-0.920313	0.438249	-2.1000	0.0357
Prov. Ancona	0.377204	0.138060	2.7322	0.0063
Marche	0.404265	0.137762	2.9345	0.0033
Other regions	0.600305	0.144529	4.1535	0.0000
Foreign	0.518078	0.479181	1.0812	0.2796
En. Year 2000	1.06874	0.181022	5.9039	0.0000
En. Year 2001	0.598356	0.185622	3.2235	0.0013
En. Year 2002	0.439551	0.170317	2.5808	0.0099
En. Year 2003	0.345938	0.159681	2.1664	0.0303
En. Year 2004	0.937754	0.168929	5.5512	0.0000
En. Year 2005	0.797484	0.176789	4.5109	0.0000
En. Year 2006	0.941962	0.171298	5.4990	0.0000

Table 19: Retention Equation: Faculty of Sciences

	Coefficient	Std. Error	z-stat	p-value
Const	10.4271	2.36008	4.4181	0.0000
Gender: male	−0.00311001	0.0699084	−0.0445	0.9645
Age	−0.0750014	0.670251	−0.1119	0.9109
Age2	−0.0220279	0.110455	−0.1994	0.8419
H.S.Grade: mean dev.	0.0221786	0.00266871	8.3106	0.0000
Av. Peer Effect	−0.127779	0.0261048	−4.8948	0.0000
Classical	−0.137449	0.0942302	−1.4587	0.1447
Languages	−0.173476	0.145573	−1.1917	0.2334
Technical	−0.580132	0.0905141	−6.4093	0.0000
Pedagogical	−0.483691	0.120711	−4.0070	0.0001
Vocational	−0.887618	0.136478	−6.5037	0.0000
Other	−0.803282	0.300178	−2.6760	0.0075
Prov. Ancona	0.261935	0.100822	2.5980	0.0094
Marche	0.184856	0.101752	1.8167	0.0693
Other Regions	0.336950	0.109805	3.0686	0.0022
Foreign	0.572895	0.352837	1.6237	0.1044
En. Year 2000	0.436256	0.137628	3.1698	0.0015
En. Year 2001	0.342731	0.139295	2.4605	0.0139
En. Year 2002	0.443300	0.126302	3.5099	0.0004
En. Year 2003	0.402965	0.121052	3.3288	0.0009
En. Year 2004	0.254492	0.126511	2.0116	0.0443
En. Year 2005	−0.0682184	0.129761	−0.5257	0.5991
En. Year 2006	−0.0581891	0.126506	−0.4600	0.6455

Table 20: Standard Deviation and Correlation Coefficient: Faculty of Sciences

	Coefficient	Std. Error	z-stat	p-value
σ_ε	1.60935	0.0325719	49.4092	0.0000
ρ	0.833593	0.0133118	62.6206	0.0000

Table 21: Conditional Mean Equation: Faculty of Sciences

	Coefficient	Std. Error	z-stat	p-value
α_0	0.517967	0.0116226	44.5655	0.0000
Const	-0.671322	1.81304	-0.3703	0.7112
Gender: male	-0.0164377	0.0542525	-0.3030	0.7619
Age	0.719972	0.540079	1.3331	0.1825
Age2	-0.106713	0.0896613	-1.1902	0.2340
H.S.Grade: mean dev.	-0.00856863	0.00217952	-3.9314	0.0001
Av. Peer Effect	-0.00849647	0.0201942	-0.4207	0.6739
Classical	0.0908035	0.0785542	1.1559	0.2477
Languages	-0.0189456	0.102477	-0.1849	0.8533
Technical	0.0597037	0.0696250	0.8575	0.3912
Pedagogical	-0.00205109	0.0907174	-0.0226	0.9820
Vocational	0.0129272	0.0984135	0.1314	0.8955
Other	-0.326590	0.177879	-1.8360	0.0664
Prov. Ancona	0.0665559	0.0805919	0.8258	0.4089
Marche	-0.0245398	0.0822817	-0.2982	0.7655
Other Regions	0.0260117	0.0898211	0.2896	0.7721
Foreign	0.304547	0.296433	1.0274	0.3042
En. Year 2000	-0.117314	0.106512	-1.1014	0.2707
En. Year 2001	0.0328019	0.100212	0.3273	0.7434
En. Year 2002	0.215627	0.0887500	2.4296	0.0151
En. Year 2003	0.223780	0.0904903	2.4730	0.0134
En. Year 2004	-0.231234	0.0966290	-2.3930	0.0167
En. Year 2005	-0.481289	0.0948624	-5.0735	0.0000
En. Year 2006	-0.546094	0.0953996	-5.7243	0.0000