The determinants of the academic outcome: an Bayesian approach using a sample of economics students from the University of Brasilia, Brazil

Ferreira Lima, Luis Cristovao

University of Brasilia, University of Helsinki

November 2012

The Determinants of the Academic Outcome: an Bayesian Approach Using a Sample of Economics Students from the University of Brasília, Brazil.

Author: Luis Cristovao Ferreira Lima*

Abstract

Using a survey conduct with 240 Economics students of the University of Brasília in August, 2011, this paper explores the determinants of the academic outcome, measured as the Gross Point Average of the University. The econometric method used to estimate is Ordinary Least Squares with Bayesian Inference. The explanatory variables include the habits of the students, such as study, frequency to classes and frequency to parties (the last one is a new approach in Brazil). Also, dummies of gender, work, type of high school and quota student were added. Study and frequency to classes turned out to be the most important determinants. The frequency to parties have not affected the Gross Point Average. The dummies had different results according to the group. There were no divergence with the major prior beliefs, with just one small exception.

Key Words: *Higher Education, Academic Outcome, Bayesian Econometrics, Affirmative Policies*

JEL Classification: *I21, I23, C11, C21, C52*

* E-mail for contact: tovaodf@hotmail.com.

1. Introduction

The discussion about the determinants of the Academic Outcome is tricky. Many papers have been written about the topic and some divergence has been found (see Shuman et al., 1985 and Michaels and Miethe, 1989). In order to contribute to this discussion, this work used a new database consisted of 240 questionnaries answered by 240 economics students of the University of Brasília, Brazil. The aim of the paper is to use Bayesian econometrics to estimate the effects of the students habits and backgrounds in their grades. The habits consist not only of time of study and frequency to classes, but also of frequency to parties and bars near the Campus, which is an approach still unexplored in Brazil.

The Bayesian Inference consists in the combination of prior beliefs and the data information. The prior beliefs can be the results of previous researchs, opinions of specialists or of the conductor of the research. As there are different priors for some of the explanatory variables, we estimate the models with two sets of priors, which consists in a prior sensitive analysis. The estimations use priors based on previous researchs, mostly by Brint and Cantwell (2008), Stinebrickner and Stinebrickner (2004), Shuman et al. (1985), Michaels and Miethe (1989) and Francis and Tannuri-Pianto (2012), among others.

In order to access the Academic Outcome, the paper also explores particularities of the Brazilian education system, such as the problematic public schools and the recent adoption of quotas for afro-descendent students in the public universities. Nevertheless, the results of the main determinants of grades are consistent with researchs worldwide, such as Brint and Cantwell (2008), Stinebrickner and Stinebrickner (2004) and Keith (1982).

The most important variable for the Gross Point Average (GPA) of a student is the amount of time spent studying outside the classroom. It has a positive and large effect on grades. In addition, the percentual of absences to classes reduces the GPA. The estimations could not find any relationship between frequency to the Campus nightlife and the academic outcome of a student. Considering the student's background, we can see that the ones who studied in a public high school performed slightly worse than the others. Also, afro-descendent quota students had lower grades. Women has an average GPA higher than men. The only result different from the prior beliefs is about the grades of people who work during their studies. On this sample, they performed the same as the average, but the other reserchs estimate a GPA lower for them.

The paper is organized as follows: Section 2 describes the database and the questionnarie used to obtain it. Section 3 shows the models which are going to be estimated and also contains a brief description of the particularities of the Brazilian education system. Section 4 presents the differents prior beliefs and the two prior sets used to estimate the models. Section 5

contains the results (Posteriors means and standard errors). And Section 6 concludes and gives suggestions for future research.

2. Database

The database used on this research is the result of questionnaires answered by 240 economics students of the University of Brasília, Brazil. It took part in the end of August, 2011. The University of Brasília is a public university located in the capital of the country, Brasília. It is a University of big proportions, with 35000 alumni and more than 5000 staff members. It is considered one of the best Universities in the country, especially in social sciences and economics. The total amount of economics alumni is 454, according to the University records. Nevertheless, the questionnaires were only answered by students between the second and seventh semesters of study. Considering only these ones, the number falls to 370. That way, the sample corresponds to 65% of the eligible population.

The questionnaires were answered in representative courses of each period of study. The questions had been designed to prevent the respondents to be identified. All the process was anonimous, which encourage best responses (Lavrakas, 2008). As the data is based in self reported information, it has probably some amount of measurement error, but it was minimized by these methods (Wilson and Zietz, 2004; Bound et al. 2001). Also, we should assume that the measurement error is random, or, in other words, it is not correlated to specific characteristics of the respondents, as gender, age, habits or income. This assumption seems to fit the data.

The first part of the questionnaire asked about the student socioeconomic background. The second part asked about the habits of this student during the previous academic period. Questions on how many hours he/she studied per week and the number of absences were the most important ones. The paper has also the intention to measure the relation between frequency to parties in the campus and the Academic Outcome. Because of it, there were questions targetting the students leisure habits. Finally, to access each one's Grade Point Average (GPA)[1], it was asked the grades in all the courses that the student has taken in the last academic period. As many students do not know their accumulated GPA (and also try to inflate it, according to Lopus and Maxwell, 1994), it was better to ask their grades on the previous period and calculate the GPA based only on those

1 The Gross Point Avarage of the University of Brasília is calculated by the formula:

$$GPA = 20 * [1-(0.6*NDcom + NDop)/NC] * sum (G * Cr)/ sum (Cr)$$

where NDcom indicates the number of compulsory courses dropped. Ndop indicates the number of optative courses dropped. NC is the number of courses the student is registered, including the dropped ones. G is the grade in an integer scale between 0 and 5. Finally, Cr is the number of credits of each course. All the expression is multiplied by 20 in order to be in a scale from 0 to 100.

grades. This methodology allows for smaller measurement error, as long as the time lag is shorter (Bradburn and Sudman, 1973).

3. Specification of the Models

In order to estimate the effects of different habits of the students in the academic outcome, it is needed to formulate models that establish a causal relationship between habits and the Gross Point Average (GPA). A simple model including only the habits would be:

Model 1:

$$GPA = \beta 0 + \beta 1 * Academic\ Center + \beta 2 * Campus\ Nightlife + \beta 3 * Study + \beta 4 * Absence + \varepsilon$$

In which β0 is the intercept, *Academic Center* indicates the number of times per month the student has been in the Academic Center of the course[2], *Campus Nightlife* measures the number of times per month the student has been to the Campus to have parties, happy hour's and drink in bars next to the University buildings (the variable is actually a sum of these three). *Study* indicates the hours per week of study (outside the classroom). *Absence* is the percentage of absences that a student had in all the courses (on average). The error term is given by ε, which has a normal distribution and it is supposed to be independent of the explanatory variables[3].

The model above can be misleading because it does not have all the necessary explanatory variables. So it can have omitted variables which bias the estimation. A more complete model with dummies variables would be as follow:

Model 2:

$$GPA = \beta 0 + \beta 1 * Academic\ Center + \beta 2 * Campus\ Nightlife + \beta 3 * Study + \beta 4 * Absence + \beta 5 * Female + \beta 6 * Public\ High\ School + \beta 7 * Work + \beta 8 * Afro\text{-}descendent\ Quota\ Student + \varepsilon$$

In which *Female* is equal to one when the student is a woman. *Public High School* indicates if the student had been to a public high school[4] (equal to one if true and zero otherwise).

2 In spite of the name, the Academic Center is not a place to study. It is place of socialization inside the Campus. There, students can talk, play games and watch television. It is a well spread habit in the public Brazilian universities and it can be somehow compared to the American fraternities, but the division is only based on the major of the student. Usually, every degree has this kind of space for it's own purposes.

3 The classical tests had not indicate heteroskedasticity (Breusch-Pagan and White test). Because of it, we can estimate using linear regression without the need of a covariance matrix for correction.

4 The Brazilian public education system is still deficient in many aspects. The public primary, secondary and high school are not good, but the public universities are the best universities in the country. Because of it, people from the middle or higher classes usually enroll their children in private schools (which are much better than the public

Work is a dummy with the positive value for students who also work. And *Afro-descendent Quota Student* is equal to one when the person had entered the University through the quota system for afro-descendent Brazilians[5].

The Model 2 can control for different kinds of students in the sample and estimate the *ceteris paribus* effects of their caractheristics[6]. Nevertheless, it is impossible to figure out the hability of an individual. This would be an important explanatory variable for the Academic Outcome. But, the absence of this variable is not a problem as long as this intrinsic hability is not correlationated with any other variable (in that case, it would be just part of the error - an unknown parameter). In Stinebrickner and Stinebrickner (2004), the authors developed a good argument for the exogenity of this variable. The greatest concern is about the possible correlation between the hability of the individual and the amount of time the person spends studying. The common sense would suggest that more intelligent people would study more. Stinebrickner and Stinebrickner (2004) admits that it is true for part of the people with high intrinsic hability (because it would make them enjoy the act of studying), but it is not true for the average of them. The explanation come from the assumption that part of these people values leisure more. It would make them study just the minimum amount of hours to have an approvable grade. As these students are more intelligent (meaning the "marginal productivity" of one hour of study is higher for them), they would actually study less than the normal student. Then, it is impossible to conclude that a correlation between studying and intelligence actually exists.

4. Prior Elicitation

Before estimating the models, it is necessary to define the prior beliefs about each of the independent variables. As a linear regression model is being used, the likelihood, the prior and the posterior will follow a Normal-Gamma distribution. The Bayesian Inference, in this case, uses the basic statement below (from the Bayes rule):

---

ones). Then, these students are the majority in the public universities (in this sample, 84.6% studied in private schools). This variable is important to determine if there are differences in the academic outcome of people coming from public and private schools.

5 Back in 2004, the University of Brasília was one of the first in the country to implement a quota of 15% of it's places to black students (20% in the first public contest of the year and 10 % in the second). It is considered an affirmative policy with the objective to diminish the disparities between races in Brazil. In 2012, the Congress approved a law disseminating this action to all the federal public universities of the country. A discussion about the effectivity and problems of this policy can be found in Francis and Tannuri-Pianto (2012) and Velloso (2006).

6 A model with more controls was also estimated (such as age and the number of courses taken in the semester), but those variables turned out to be not important. Also, when dummies for different incomes were added we can not find correlation between incomes and grades, which is in accordance with White (1982). But, as 86% of the students are on the top decil of the Brazilian income distribution (Andifes, 2011), we can deduce that income is a determinant of being admitted in the University of Brasilia. Alternative models can be presented upon request.

*p(β,h|y) is equivalent to p(y|β,h) * p(β,h)*

Where *p(β,h|y)* is the posterior, *p(y|β,h)* is the likelihood of the data and *p(β,h)* is the prior. The posterior is a product of the likelihood and the prior. In the linear regression model, the likelihood follows a Normal-Gamma distribution and the prior is also designed to follow a Normal-Gamma distribution. Because of it, The posterior will also be Normal-Gamma (Koop, 2003).

We have to determine the values of β and h to use in the prior. For each β parameter, we must attribute a value (mean) and the degree of certainty of it (variance). The estimation will use different priors according to previous studies and theories. The parameter h indicates the error precision of the estimation. It is set as 0.1.

First, the parameters which do not have divergence in the previous studies are *Absence, Public High School, Work* and *Afro-Descendent Quota Student*. Shuman et al. (1985), Stinebrickner and Stinebrickner (2004), Stinebrickner and Stinebrickner (2008), Brint and Cantwell (2008), Francis and Tannuri-Pianto (2012), Durand and Rau (2000) and Michaels and Miethe (1989) all agree that the number of absences has a large negative effect on grades. They also agree that students who work have a slightly smaller GPA. In a huge survey in the University of Brasília, Francis and Tannuri-Pianto (2012) showed that the ones who studied in Public High Schools have a GPA below the average, but statistically non-significant (using classic inference). Velloso (2006) and Francis and Tannuri-Pianto (2012) found that Afro-descendents Quota Students have a GPA smaller than the others, and it is statistically significant (but the difference is not big, making the authors argue that this policy is effective).

But, there are divergences about the signal and magnitude of the other variables. The frequency to the Academic Center and to the Campus Nightlife would have a negative impact on grades, according to Dowdall et al. (1998) and they are quite sure of it. On the other hand, the surveys made by Brint and Cantwell (2008) in the University of California and Stinebrickner and Stinebrickner (2008) in the Berea College, Canada, have not found any relationship between social nightlife and GPA, and the size of the sample was above 5000 students in each research.

Also, the findings for the effect of hours of study in grades are ambiguous. Shuman et al. (1985) made several researches in the University of Michigan from 1975 to 1984 and could not find any relationship between study and GPA. But, all the other papers have presented a positive and significant effect between these two variables.

Finally, the difference in GPA according to genders are not the same in all the researches. In Kenneth and Melaugh (2009), the authors used data from more than 35000 principles of economics students from 1967 to 2007 and concluded that men performed better than women in the course “Principles of Economics” of the University of Virginia. This finding is exactly the opposite of the majority. The other papers concluded that women have better grades than men. For

the University of Brasília, the last findings also hold. Francis and Tannuri-Pianto (2012) and Velloso (2006) found that women have a better performance.

The prior parameters are in Table 1.

Table 1 – Prior Parameters of β

Variable	Mean 1	Variance 1	Mean 2	Variance 2
Academic Center	0	0.4	-0.2	0.3
Campus Nightlife	0	0.15	-0.3	0.2
Study	0.6	0.25	0	0.5
Absence	-0.25	0.15	-0.25	0.15
Female	1.	0.6	-2	2
Public High School	-2	1	-2	1
Work	-1.5	1	-1.5	1
Afro-Descendent Quota Student	-4.5	2	-4.5	2
Constant	70	3	70	3

Source: Data and previous studies.

The second and third columns indicate the means and variance of the first set of priors. These ones are the most feasible, considering previous studies and the data. The prior of the constant was set close to the average GPA, which is 71. It means that we expect the other explanatory variables to affect the academic outcome toward the mean. Then, we expect that the frequency to the Academic Center and Campus Nightlife will not affect the GPA. Study and Absence are expected to have a huge impact on academic performance. Women should have better grades, but people who work, went to a public high school or is an afro-descendent quota student are expect to have smaller GPAs.

The fourth and fifth columns show the priors which fit with some previous studies but do not seem to be in accordance with the data obtained. In the next session, we will show the results and explain why the estimations with the first list of priors are better.

5. Results

On the Table 2 below we present the posterior means of the estimation. It is the result of the combination of the prior information and the data information (which dominates).

Table 2 – Posterior Means and Standard Errors*

Variable	Model 1 with the first Priors	Model 1 with the second Priors	Model 2 with the first Priors	Model 2 with the second Priors
Academic Center	0.061 (0.095)	0.062 (0.095)	0.023 (0.097)	0.025 (0.097)
Campus Nightlife	-0.029 (0.177)	-0.066 (0.185)	-0.026 (0.175)	-0.064 (0.183)
Study	0.594 (0.128)	0.588 (0.130)	0.624 (0.125)	0.616 (0.127
Absence	-0.218 (0.047)	-0.216 (0.047)	-0.199 (0.047)	-0.198 (0.047)
Female	-----	-----	3.567 (1.076)	3.148 (1.411)
Public High School	-----	-----	-1.892 (1.604)	-1.956 (1.611)
Work	-----	-----	0.011 (1.340)	0.040 (1.345)
Afro-Descendent Quota Student	-----	-----	-6.539 (2.771)	-6.540 (2.781)
Constant	70.305 (1.585)	70.426 (1.599)	69.304 (1.702)	69.605 (1.765)

***** *Standard errors in parentheses*

As said before, the first set of priors are more suitable to the data. Also, the model with more variables can explain considerably more about the Academic Outcome. The Bayes Factor (Posterior Odds Ratio when the probability of the models are set as equal) give us the likelihood of a model in comparison to other model. A Bayes Factor bigger than one gives support to the first model and this support can be conclusive if the number is bigger than ten[7]. As shown in the Table 3 below, the model with the first set of priors and the greatest number of variables is the most likely to occur among the models available.

7 Actually, the result of the Bayes Factor gives us how many times the first model is more likely to occur than the second model. A number above ten implies that the first model is more than times more probable.

Table 3 – Bayes Factors between models*

	Model 1 with the first Priors	Model 2 with the first Priors
Model 1 with the first Priors	------	43.31647
Model 1 with the second Priors	2.425328	105.0566
Model 2 with the second Priors	------	10.38739

Source: Own calculations.
** The model tested is the one specified on the top of each column.*

We conclude that the best model to interpret the results is the Model 2 with the first Priors. It is possible to see by the results in Table 2 that the data dominates the misleading second Priors. The Posterior means are almost the same with the two different lists of Priors.

The results show that the time studying is the strongest predictor of the Academic Outcome, opposing the work of Shuman et al. (1985), but in accordance to more recent studies. One additional hour of study per week indicates an increase in GPA of 0.624. Also, absence to classes play a role on the student's grades. One percent more of faults represents a decrease of 0.199 on GPA, which is similar to previous papers. On the other hand, the party habits seem not to affect the student's Academic Outcome. The variables *Academic Center* and *Campus Nightlife* have means close to zero. Even when negative priors were used, these variables continue to do not have a relationship with GPA (see Table 3, column 5).

The females performed better than males. They have a GPA 3.567 higher. This result is different from Kenneth (2009), but in accordance to the others studies. The students who studied in a public high school presented a GPA below the average (-1.892), showing that some background differences still persist even in Universities. The data shows us that there is no difference between the GPA of people who work and people who do not. This result is contrary to the literature and there is no easy explanation for it[8]. Afro-descendent quota students have worse grades, even more than what was predicted in previous studies. This result should be analysed carefully. It is impossible to tell for sure if this is the true mean for this group. First, there was only 12 of these alumni in the sample, which is quite misleading[9]. The result is lower than the prior belief (based on Velloso, 2006, and Francis and Tannuri-Pianto, 2012). Perhaps, that is a particularity of the

8 We could guess that people who work have, somehow, more abilities than the average. So, despite spending time working, they can have the same results of the others, *ceteris paribus* (if we set all the variables at the same value).

9 The expected number of Afro-descendent Quota students in the sample would be 36 (15% of 240). But, it was actually 12 (5%). There are three non-excludent possibilities: first: many quota students would not report it. Second: quota students could skip more classes than the others, making it difficult to get their data. Third: quota students can have a higher rate of abandonment in the University of Brasília (and they actually do, according to Francis and Tannuri-Pianto, 2012). That way, it is possible that their grades in the Economics courses are even lower, but we do not have enough elements to make any affirmation. This way, it would not been ethical to use these results to produce arguments against this affirmative policy.

economics students, but we can not conclude much from this result.

Using these estimations, the typical economics student[10] will have a predict GPA of about 72, close to the data average. If we consider a white woman who works and studies 20 hours per week and do not misses classes or go to parties, her expected GPA would be around 85.3, much higher than the previous student. This difference is due to the effort related to time studying and class atendance, but we can not say that it is because of the frequency to the Campus Nightlife.

6. Conclusion

The estimation of the determinants of the Academic Outcome among the students of economics in the University of Brasília showed that effort means achievement, on average. The most important variables were time studyiing and frequency to classes. Both of them measure the compromise of the student to the University.

Because the role of effort, we could expect that leisure should be mutually exclusive to the time studying. But, we observe a different path. The time spent partying and socializing do not affect the GPA of the students, which is in accordance to Bint and Cantwell (2008) and Stinebrickner and Stinebrickener (2004).

The study also analized the differences on academic achievement between groups. We figured out that women performed better, with a GPA 3.5 higher, ceteris paribus. The alumni who had studied in a public high school had slightly worse grades. Also, the afro-descendents quota students had a GPA 6.5 below the average. The students who work had the same grades of the others, which is a peculiar result.

This research tried to understand how the habits and backgrounds of a student affect his academic achievement on the University. The results showed how unequal the Brazilian society is. The economics major is almost totally occupied by students of a extremely high income. Although, affirmative policies were developed in order to diminish the income and racial differences in Universities and in the country as whole. But, the effects of these policies are still ambiguos. More studies should be done in order to access measures and ideas to develop the Brazilian education system. Only a massive investment on this sector could make Brazil a high develop country with similar opportunities for the whole population. The time for it is running.

10 The typical economics student in the University of Brasília is a man who studies 8 hours per week, misses 12% of the classes, does not work, studied in a private high school, is not black, goes 7 times per month to the Academic Center and 4 times to parties on the Campus.

7. Bibliography

ANDIFES. Perfil Socioeconômico e Cultural dos Estudantes de Graduação das Universidades Federais Brasileiras (Cultural and Socioeconomical characteristics of the undergraduate students of the Federal Brazilian Universities). July, 2011.

BANDURA, Albert; MARTINEZ-PONS, Manuel; ZIMMERMAN, Barry J. Self-Motivation for Academic Attainment: The Role of Self-Efficacy Beliefs and Personal Goal Setting. *American Educational Research Journal*, Vol.29, No.3, p.663-676. 1992.

BRADBURN, Norman M.; SUDMAN, Seymour. Effects of Time and Memory Factors on Response in Surveys. *Journal of the American Statistical Association*, Vol.68, No. 344, 805-815, 1973.

BOUND, John; BROWN, Charles; MATHIOWETZ, Nancy. Measurement Error in Survey Data. *Handbook of Econometrics*, Vol. 5, No.1, 2001.

BRINT, Steven; CANTWELL, Alisson M. Undergraduate Time Use and Academic Outcomes. *Research and Occasional Paper Series of the Center for Studies in Higher Education*, University of California, Berkeley, 2008.

DOWDALL, George W.; GLEDHILL-HOYT, Jeana; LEE, Hang; MAENNER, Gretchen; WECHSLER, Henry. Changes in Binge Drinking and Related Problems Among American College Students Between 1993 and 1997. *College Health*, Vol.47, No.1, 1998.

DURAND, Ann; RAU, William. The Academic Ethic and College Grades: Does Hard Work Help Students to “Make the Grade"? *Sociology of Education*, Vol.73, No.1, p.19-38, 2000.

FRANCIS, Andrew M.; TANNURI-PIANTO, Maria. Using Brazil's Racial Continuum to Examine the Short-Term of Affirmative Action in Higher Education. *Journal of Human Resources*. Vol. 43, No. 3, p.754-784. 2012.

IBGE; Brazilian Institute of Geography and Statistics. 2010 Census; Brasilia, 2011.

KEITH, Timothy Z. Time Spent on Homework and High School Grades: A Large Sample Path Analysis. *Journal of Educational Psychology*, Vol.74, No.2, p.248-253, 1982.

KENNETH, Elzinga; Daniel, MELAUGH. 35,000 Principles of Economics Students: Some Lessons Learned. *Southern Economic Journal,* Vol. 76, No. 1, pp. 32-46. 2009.

KOOP, Gary. Bayesian Econometrics. England. Wiley Publications. 2003.

LAVRAKAS, Paul. Encyclopedia of Survey Research Methods. United States of America. SAGE Publications. 2008.

LOPUS, Jane S.; MAXWELL, Nan L. The Lake Wobegon Effect in Student Self-Reported Data. *The American Economic Review*, Vol.84, No.2, p.201-205, 1994.

MICHAELS, James W.; MIETHE, Terance D. Academic Effort and College Grades. *Social Forces*, Vol.68, No.1, p.309-319, 1989.

OECD. Education at a Glance 2011. OECD Publications. September, 2011.

SHUMAN, Howard; ETHERIDGE, Barbara; OLSON, Camile; WALSH, Edward. Effort and Reward: The Assumption that College Grades are affected by Quantity of Study. *Social Forces,* Vol.63, No.4, p.945-966, 1985.

STINEBRICKNER, Ralph; STINEBRICKNER, Todd R. Time-Use and College Outcomes. *Journal of Econometrics*, Vol. 121, No.1, p.243-269, 2004.

STINEBRICKNER, Ralph; STINEBRICKNER, Todd R. The Causal Effect of Studying on Academic Performance. *The B.E. Journal of Economic Analysis & Policy*, Vol.8, No.1, p.1-53, 2008.

VELLOSO, Jacques. Curso e Concurso: Rendimento na Universidade e Desempenho em um Vestibular com Cotas da UnB. NESUB e Faculdade de Educação da UnB. 2006.

WHITE, Karl R. The Relation Between Socioeconomic Status and Academic Achievement. *Psychological Bulletin*, Vol.91, No.3, p.461-481, 1982.

WILSON, Mark L.; ZIETZ, Joachim. Systematic Bias in Student Self-Reported Data. *Journal for Economic Education*, Vol.4, No.4, p.13-19, 2004.

WOOLDRIDGE, Jeffrey M. Econometric Analysis of Cross Section and Panel Data. First Edition. MIT Press Books. 2001.