The Influence of Family Background and Individual Characteristics on Entrance Tests Scores of Brazilian University Students

Juliana Guimarães*
Breno Sampaio†

Revised in July 2007

Abstract

This article examines the factors that influence university students performance on the entrance test at Universidade Federal de Pernambuco, Brazil. Particular attention is paid to the importance of student background and educational resources. The results suggest that parents education and study environment are key determinants of students achievements. Religion, race and gender do paly a role on their performance. Above all, the quantile regression estimation show that most of these factors vary across the conditional score distribution, showing that one isolated factor can have a different impact across students.

Key words: Academic Achievement, Family Background, Quantile Regression.

Resumo

Este artigo analisa os fatores que afetam o desempenho de estudantes no vestibular da Universidade Federal de Pernambuco. Destaque é dado para a importância do background familiar e educacional. Os resultados mostram que a educação dos pais e o ambiente de estudo são determinantes para o desempenho acadêmico. Religião, raça e gênero também influenciam no desempenho. Estimações de regressão quanítlica mostram que a influência da maioria desses fatores varia ao longo da distribuição condicional da nota, mostrando que o impacto de um fator isolado tem influência distinta entre os estudantes.

Key words: Desempenho Acadêmico, Background Familiar, Regressão Quantílica.

JEL Cassification #: J01, J24, C13

*Universidade Federal de Pernambuco, Departamento de Economia-PIMES, Av.dos Economistas, S/N, Cidade Universitária, Recife-PE -Brazil, 50740-590, guimaras@fe.unl.pt.

†University of Illinois at Urbana-Champaign, Department of Economics, brenosampaio@hotmail.com. We are grateful to Tiago Cavalcanti and Yony Sampaio for their comments and suggestions. All the remaining errors are our responsibility.

1 Introduction

Over the past decades researchers have examined the key determinants of student performance on standardized achievement tests. Data from England and United States bringing information on student performance, their family background, and school characteristics had provided researchers with a true live laboratory to answer questions such as the effect of income, work hours, and school characteristics, among others, on students achievement. This area of study is still new in Brazil. There are few studies evaluating determinants of performance of university students.

In this paper we use an unique data set on students academic scores at Universidade Federal de Pernambuco (UFPE) which is the major University in Northeast of Brazil[1]. The data set brings information on their standardized entrance test scores, personal characteristics, such as age, gender, race, religion, family background, and high school attended. It allows us to estimate the key determinants of students performance on the entrance test. Using Ordinary Least Squares (OLS) and quantile regression we estimate not only the mean effect, but also the effect of these variables on different quantiles of the conditional score distribution. The conditional quantile function will give a family of functions, one for each τ, which provide a more complete characterization of the relationship between test scores and students characteristics, compared to the one given by OLS regression, which concentrates on the first conditional moments.

Gender differences in academic performance may arise for a number of reasons such as differences in the types of subject male and female students study, gender differences in individual-specific attributes that are correlated with attainment (e.g. family background, age and marital status) or differences in the type and quality of the institutions male and female students attend (Hoskins *et al* . 1997; Rudd 1984). According to Mellanby *et al* (2000), gender differences in attainment could be due to psychological and/or biological factors. Our estimated coefficient for the gender variable shows that women tend to perform worse than men. Moreover this result is stronger for those on the upper quantiles. Therefore, not only women has lower entrance test scores (ETS) compared to their male counterparts but also, this is larger for students at upper quantiles of the conditional score distribution. These findings go in the opposite direction to those on the international empirical literature. We decided to look deeper and examine female performance among different University departments. The results show that women tend to have higher grades on predominantly male departments such as the Engineering College. This also happens on the Health Department. At Human Sciences and Social Sciences women at low quantiles have lower grades than men, this changes as we move along the conditional score distribution.

The presence of children also tend to lower women grades. Not surprisingly, more educated parents tend to have a positive effect on students scores, and this is stronger for above average students. Family income also influence grades positively specially on the upper quantiles, although the fact that the student is still living with parents tend to have a negative impact on their scores. Students that work perform worse compared to those that do not and this negative effect is stronger for students at top quantiles of the conditional score distribution.

[1]Established in 1946, the Universidade Federal de Pernambuco is according to the Education Ministry the major University on the North/Northeast regions of Brazil. The courses are offered by 10 colleges of four different areas which consist of 67 Departments. The University offers 62 different undergrad courses, 17 of them are ministered at night. The University also has 108 post graduate courses, among masters, PhDs and MBAs. In 2004 the University had 25,000 students registered (20,500 were undergrads and 4,500 were graduate students) and 1,647 professors.

Students that attended public schools tend to perform worse than private school students. The longer they spend at the public school system the worse. This result is stronger for students at the upper quantiles of the score distribution, compared to those located at the lower tail. According to our sample statistics this is a composition of worse family background and poor school quality that leave public school students in disadvantage compared to private schools ones. Notice that despite the grade disadvantage of public school students when we controlled for school attended we notice that among the top five schools in our sample four are private, however the very top one is a public school.

There is no consensus concerning the effect of hours worked during school on students current and future performance (Steinbricker and Steinbricker, 2003). Here we found that the higher the number of working hours the lower the student performance on the test on average. Every additional hour lower the students score on 3.7 percentage points on average. The quantile results show a different picture.

The structure of this paper is as follows. Section 2 presents the methodology. Section 3 describes the data. Section 4 present the least square and quantile regression estimates and Section 5 concludes.

2 Methodology

This paper uses quantile regression to estimate the relationship between students achievement in entrance tests to the University and their demographic characteristics. We estimate this function using both OLS techniques and quantile regression [2] . Notice that Y_i is the dependent variable, entrance test scores and X_i a matrix of explanatory variables including personal characteristics, family background, school attended among others. In the same way that $\hat{\beta}_{\mathrm{OLS}}$ minimizes the sum of the loss function, $(log(Y_i) - X_i'\beta)^2$, $\hat{\beta}(\tau)$ minimizes the sum of the following linear loss function $\rho_\tau(log(Y_i) - X_i'\beta)$. [3] Thus the τ-th conditional quantile function is given by

$$Q_{\mathrm{log(Y_i)}}(\tau|X) = X'\beta(\tau) \tag{1}$$

The conditional quantile function will give a family of functions, one for each τ, which provide a more complete characterization of the relationship between $\log(Y_i)$ and X compared to the one given by OLS regression, which concentrates on the first conditional moments (Arias, Hallock and Sosa (2001)) . In addition, Koenker and Portnoy (1996, pp. 36-42) show that the quantile functions have, in general, the same robustness properties to outlying observations as the ordinary τ-th sample quantiles. These robustness properties are very important when the distribution of the disturbance term deviates from the Gaussian distribution. With the quantile model the entrance test scores can be influenced by personal characteristics in different ways at different parts of the distribution.

Finally, the slope parameters of the family of estimated quantile functions provides a way to test for the presence of heteroskedasticity in the model (Koenker and Bassett, 1982). For instance, if

[2] We do not discuss quantile regression in detail. Instead we will just comment on some important properties of this approach, which are useful for this study. We suggest the works of **?**, **?** and **?** as comprehensive sources of how to understand quantile regression.

[3] $\rho_\tau(u) = u\tau - uI(u \leq 0)$ where $I(u \leq 0)$ is an indicator function.

some slope coefficients are changing with τ then this is indicative of some form of heteroskedasticity. Therefore, in the quantile regression technique we can address the heterogeneity of the unobservable effects in an informative and constructive way. Formally, Koenker and Bassett (1982) propose a Wald-type statistic to test if the slope parameters are equivalents for different quantiles.[4]

Quantile regression has been used extensively in the economics literature to analyze gender wage differentials, returns to education and income inequality, and recently to examine students achievements [5].

3 The Data

In this paper we use an unique data set on students entrance test scores at Universidade Federal de Pernambuco (UFPE) which is the major University in the Northeast of Brazil. The students were taking the entrance test (*vestibular*) in 2005. University student record data are very rich in characteristics of individuals, the colleges they are applying for and their previous school. Our dependent variable is their scores on the entrance test. The explanatory variables can be divided into, *Personal Information* such as age, gender, marital status, race, religion, number of children, parents schooling, parents employment status, family income, hours worked and *Academic History* such as school attended (so we can identify the type of school if private or public, catholic or other), if had lab classes, foreign language classes, preparation classes to the entrance test and the college they applied for among others.

3.1 Summary Statistics

Table 1 brings information on summary statistics on the key variables of interest. Data on 56,723 students who took the Entrance Test to Universidade Federal de Pernambuco in 2005 was collected. Cases with missing values of variables included in the study are omitted. This leaves 54,877 students in the sample used in the statistical analysis covering more than 90% of all students with roughly equal numbers of males and females. The main explanatory variable is the students achievement on the entrance test that ranges form 0 to 10.

Our sample consists of students which are 20 years old on average. Most of them single and still living with their parents. The majority of our students classifies themselves as white or *pardos* [6]. More than 50% of the students are catholics, 21% are protestants, 11% declared themselves as Atheists and less than 1% are Jewish.

The average monthly family income is R$1620.00 (around 4.5 minimum wages). Notice however that the income distribution across families is very unequal as shown by the standard deviation. Parents have around 11 years of education. Almost 60% of students fathers are working while 50% of their mothers have paid jobs. On average 27% of the students are working around 1.6 hours per day.

[4]See Koenker and Bassett (1978) and Koenker and Portnoy (1996) for more details.

[5]See for instance Eide & Showalter (1998), Ng & Pinto (2003), Bassett *et al* (2002), Kremer & Levy (2003), Bassett *et al*(2002), Smith and Naylor (2005)and Birch and Miller (2006),Guimares(2007).

[6]Due to interbreeding of races (blacks and whites, natives and whites and blacks and natives) the individual classifies himself as brown or *pardos*

Table 1: Summary Statistics

Variables	**Mean**	**Stand. Dev.**
Observations	54,877	
Entrance Test Score	4.357	1.385
Age	20.538	5.402
Female	0.558	0.497
Married	0.068	0.252
Father Education	11.607	4.403
Mother Education	11.780	4.457
Working Father	0.586	0.492
Working Mother	0.499	0.500
Family Income	1,620.454	2,072.072
Working Students	0.278	0.448
Living with Parents	0.803	0.398
Whites	0.434	0.496
Asians	0.048	0.213
Natives	0.018	0.132
Pardos	0.384	0.486
Blacks	0.090	0.287
Catholics	0.545	0.498
Protestants	0.218	0.413
Jewish	0.002	0.041
Afro-Religion	0.006	0.077
Other Religions	0.091	0.287
Atheist	0.110	0.313
Hours Worked	1.690	2.921
Public School Students	0.480	0.499
Private School Students	0.519	0.499
Internet User	0.347	0.476
Reading	0.287	0.452
Lab. Classes	0.360	0.480
Foreign Language	0.042	0.201
Private Classes	0.400	0.490
Supletivo	0.034	0.182

Note: sample averages and sample standard errors

When it comes to the education system, 52% of the sample come from private schools while 48% of them studied on the public education system. Public School students have lower scores compared to private ones, 3.9 against 4.7. Students were queried about the access of educational resources. In our sample 34% of the students have access to internet, 36% have additional lab classes and only 4% of the students have extra foreign languages classes.

In Brazil, the Education Ministry offers an alternative education method for those individuals who have either drop off or did not have the chance to go to school when they should have. Those are individuals that have usually a large distortion age/grade, sometimes even illiterate adults. This alternative method is called *Supletivo* (Supplementary) and offers short-term courses with a condensed material for different grades. The students can have for instance, middle school diploma in a one year course. In our data set we have the information if the student graduated from Supletivo or not, 3% of our sample got a high-school Supletivo degree. Next section analyzes the least square results.

4 Results

4.1 Least Squares Estimates

In this section we analyze the OLS and the quantile regression estimates from the quantile function $Q_y(\tau|x) = x'\beta(\tau)$ given by $\beta_i(\tau)$, i=1,...,k and $\tau\epsilon(0,1)$. Note that rigourously, this analysis should be cared out with the dependent variable transformed to reflect the nature of the weighted average score, which is bounded by 0 and 10. [7] Observe that the the OLS is not equivariant to monotonic transformation and therefore we cannot recover from the transformed model the true effect of each independent variable on the ETS. Since the estimations with the transformed variable however, did not differ widely from the ones analyzed in this section, and they are easier to interpret, we just report the results for the untransformed model.

Most of our least squares estimates are statistically significant. Female students, on average, have grades that are 2.2 percentage points lower than their male counterparts (Table 3). This result was not consistent with the findings in the empirical literature on the determinants of academic success (see for instance, Birch and Miller, 2006). We then controlled for the college they were applying to. On this second specification it turns out that women still performed worse than men in most of the colleges. At the engineering College however the indicator was not statistically significant. Women performed worse at Medical School and Human Sciences. The presence of children lower women grades in 1.5 percentage points. These results change for different quantiles

[7] The transformation ensure that estimations will not exceed 10 or far bellow zero. The transformation of the dependent variable goes as follows:

$$y_{transformed} = log(\frac{y}{10-y}) \tag{2}$$

Marginal effects may be calculated from the estimates obtained with this dependent variable using:

$$\frac{\partial y}{\partial x}\frac{1}{y} = \beta(10-y) \tag{3}$$

These partial effects are usually evaluated at the mean value of the dependent variable $\bar{y}$.

of the conditional scores distribution. Notice that age affects students scores negatively on average, which means that the older the student the worse he/she will score on average.

Not surprisingly, a better family background is a key determinant of student performance. Controls for parents education show that parents with higher schooling have a positive impact on student scores. This is also the case for the educational resources variables such as access to internet, extra laboratory classes, preparation and foreign language classes that are all good proxies for a better study environment. Students that have access to internet increase their grade by 2.5 percentage points on average, while students that have lab classes increase their grades by 1.7 percentage points. The strongest effect goes to students that have foreign language classes who increase their scores by 5.3 percentage points. Students that declare having reading habits also had their scores increased by 2.3 points on average. Also, students that had extra private classes for the entrance tests increased their scores by 2.8 percentage points.

All these results are highly connected to the fact that, students that come from higher income families tend to perform better, given that richer parents may provide better conditions and thus a better study environment. Notice however that the fact the family has a higher income will not necessarily mean that it will give the student a better study environment, but at least it means that it has the conditions to do so.

Students who attended public schools have lower scores compared to private school students. Moreover the longer the student remained at the public school system the worst he/she will perform. Control variables for each school show that despite the fact that on a performance scale the top five schools are private ones, the first on the rank is a public school.

The lower performance of public school students reflects in fact their socio-economic status. According to Sampaio e Guimarães (2007), the average family income of private schools students is three times higher compared to those who attended public school (see Table 2). Private school parents have on average 3 more years of schooling than parents of public school students. While 15% of public school fathers were unemployed by the time of the test, only 6% of privates school fathers did not have a paid job. Students who attended public school also have a higher working load (2.5 times greater) compared to private school students. The percentage of female students that have children is also higher among public school students (4.4% against 1.9%). As we can see from Figure 1, public school students score distribution is located to the left of private school students, showing that public school students perform worse for all quantiles of the empirical score distribution.

Students who graduated with a *supletivo* degree had grades 4.1 percentage points lower on average. Moreover, we controlled for the fact that the student is taking the entrance test just for experience [8], it turns out that these students perform worse than the ones taking the test for real. Notice also that students that took the test more than once tend to perform better on average.

The fact the father is working has a positive and significant effect on student performance. Working mothers however, do not impact the same way on students scores.

Asians, Native Brazilians and Blacks tend to perform worse than white students on average. When it comes to religious differences, Atheists, Afro-religion, Jewish, Protestants and those who

[8] Some students decide to take the test on their junior high school year just to have a feeling on how they will perform on their senior year.

Table 2: Summary Statistics

Variables	**Private**	**Public**
N =	29,480	27,242
Score	4.720	3.967
	(1.406)	(1.250)
Age	19.330	21.885
	(4.197)	(6.215)
Number of Kids/Female Student	1.445	1.605
	(0.649)	(0.702)
Father Education	13.520	9.713
	(3.902)	(4.036)
Mother Education	13.715	9.878
	(3.953)	(4.092)
Hours Worked	1.110	2.318
	(2.432)	(3.257)
Income	2,421.407	837.932
	(2,496.164)	(1,077.832)

Note: [1]Standard Deviation in parenthesis.

declared having other creeds, all performed better than catholics on average (3.7, 5.9, 5.1, 0.64 and 2.4 percentage points, respectively). There is no consensus concerning the effect of hours worked during school on students current and future performance (Stinebrickner & Stinebrickner, 2003). Here we found that the higher the number of working hours the lower the student performance on the test on average. Every additional hour lower the students score on 0.38 percentage points on average.

Some of the above results differ for other parts of the distribution, different from the mean. The quantile regression estimates bellow will show the major differences.

4.2 Quantile Regression Estimates

In this subsection we analyze the quantile regression estimates from the quantile function $Q_y(\tau|x) = x'\beta(\tau)$ given by $\beta_i(\tau)$, i=1,...,k and $\tau\epsilon(0,1)$. The results are displayed in Figure 3 and 4. The plots show the quantile regression (QR) estimates as well as the 95% confidence intervals. The least square estimates are presented as the dotted horizontal line.

These plots tell a different story from the least square results. The least square results showed for instance that students that come from higher income families tend to perform better on average. The quantile regression results show that this is stronger for students in the upper quantile of the conditional score distribution. That is, students that perform better tend to take higher advantage from the family better economic condition.

Female students on average perform worse than men. The quantile estimates show that despite the fact that female students score worse compared to their male counterparts this result is not

Table 3: Ordinary Least Square (OLS) Regressions

Variables	**(1)**	**(2)**
Married	0.111* (0.031)	0.179* (0.030)
Female	-0.230* (0.012)	
· Female x Social Sciences		-0.098* (0.031)
· Female x Human Sciences		-0.395* (0.027)
· Female x Engineering		-0.177 (0.048)
· Female x Health		-0.521* (0.031)
· Social Sciences		0.165* (0.024)
· Human Sciences		0.454* (0.024)
· Engineering		0.474* (0.023)
· Health		0.807* (0.028)
Number of Kids/Student	0.058* (0.019)	
· Number of Kids/Female Student		-0.111* (0.020)
Age	-0.046* (0.002)	-0.038* (0.002)
Father Education	0.019* (0.002)	0.019* (0.002)
Mother Education	0.018* (0.002)	0.018* (0.002)
Family Income	8.26E-5* (0.000)	7.80E-5* (0.000)
Living with Family	-0.164* (0.017)	-0.133* (0.017)
Working Father	0.039* (0.012)	0.036* (0.012)
Working Mother	0.020*** (0.012)	0.017 (0.012)
Afro-Religion	0.637* (0.103)	0.597* (0.099)
Protestant	0.058* (0.014)	0.064* (0.014)
Jewish	0.550* (0.155)	0.516* (0.149)
Atheist	0.347* (0.019)	0.379* (0.019)
Other Religion	0.230* (0.021)	0.244* (0.021)
Asian	-0.069* (0.026)	-0.071* (0.026)
Indian	-0.327* (0.042)	-0.310* (0.041)
Mulatto	0.029* (0.013)	0.027** (0.012)
Black	-0.088* (0.020)	-0.072* (0.026)
Years in Pub. School	-0.019* (0.002)	-0.018* (0.002)
Supletivo	-0.417* (0.034)	-0.397* (0.034)
Hours Worked	-0.043* (0.002)	-0.038* (0.002)
Internet User	0.269* (0.015)	0.260* (0.014)
Lab. Classes	0.187* (0.013)	0.171* (0.013)
Foreign Language	0.535* (0.033)	0.534* (0.032)
Private Classes	0.317* (0.012)	0.280* (0.012)
Reading	0.252* (0.013)	0.230* (0.013)
Tests Taken	0.420* (0.007)	0.410* (0.007)
Vest. Exp.	-0.218* (0.032)	-0.254* (0.032)
Intercept	3.872* (0.050)	3.413* (0.048)
N	42.587.	42.587.

Note: *p<0.01. **p<0.05. ***p<0.10.

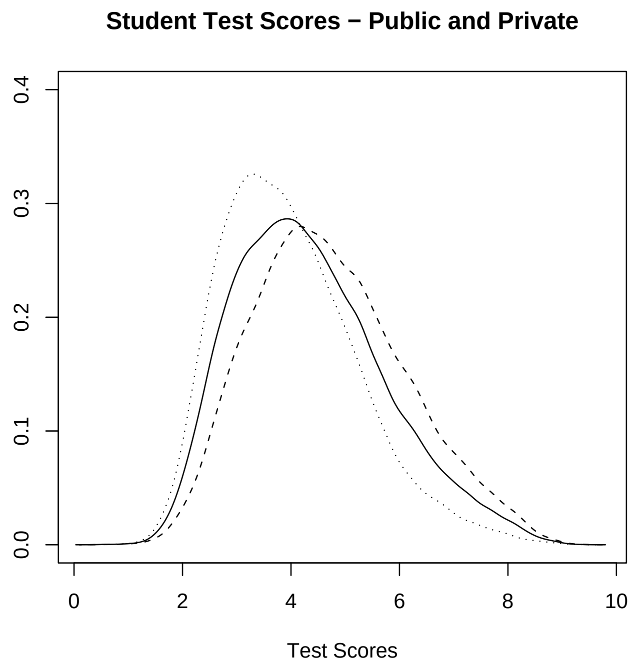

Figure 1: The above plot presents the unconditional densities estimations for student test scores in public (dotted) and private (dashed) schools compared to the density for the entire sample (solid).

monotonic across the conditional score distribution. In fact, female students at the beginning of the score distribution start with a grade advantage of 1.0 percentage point and at the upper tail this grade difference amounts 3.0 percentage points against women (Figure 2). That is, female students at higher quantiles tend to suffer from a higher grade disadvantage, compared to those at the lower quantiles of the conditional score distribution. Given that this result go against the empirical literature we decided to control for the different colleges the students were applying for. It turns out that this grade still remains for female students in some colleges such as engineering and health department. Moreover, these results change completely for different quantiles of the conditional score distribution. Females in Human Sciences for instance, start with a lower grade disadvantage at the beginning of the distribution. This scenario worsens as we move along the conditional score distribution. Both at the Social Sciences and at the Human Sciences departments, female at the upper quantiles suffer from a higher grade disadvantage at the entrance test, compared to male students.

The presence of kids not only compromise female students performance on average, but also for students at upper quantiles this effect is even stronger. That is, for female students located at upper quantiles of the conditional score distribution the presence of children has a stronger negative impact, compared to those at low quantiles.

When it comes to family background the results show that parents education, not surprisingly, affect positively student scores. Notice however that students at higher quantiles take greater advantage of more educated mothers, compared to students at lower quantiles. We may conclude that, more educated mothers provide a better study environment that contributes positively to

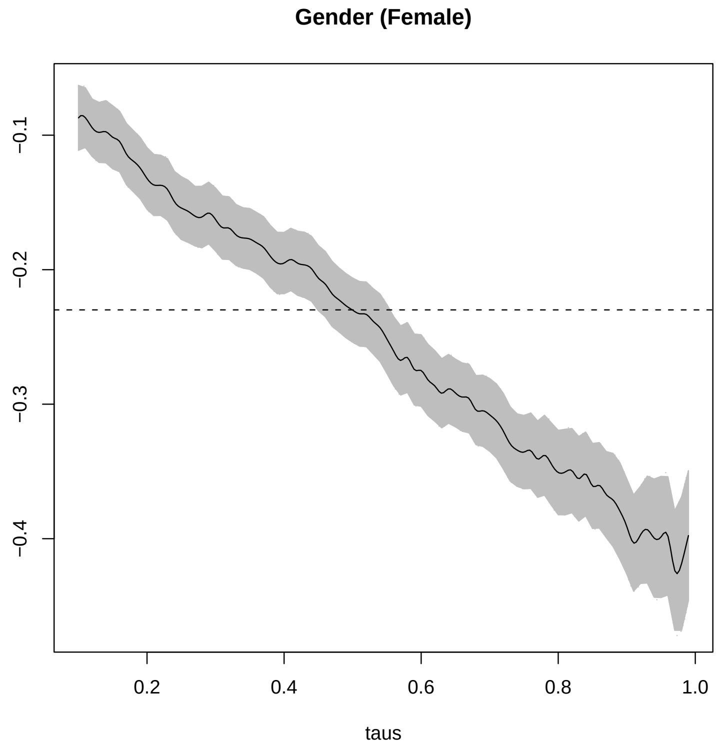

Figure 2: The above plot present the quantile regression estimates for gender (female). The solid lines are the quantile estimates and the shade the 95% confidence intervals. The dotted line presents the Ordinary Least Square estimate.

students performance. Some of our controls work as proxies for both study environment and quality of education. Students that have access to internet have higher grades compared to the ones that do not have. This grade advantage is higher for students at higher quantiles of conditional grade distribution. This is also true for students that have foreign language classes, lab classes and preparation classes for the test. Student that had private preparation classes at the lowest quantiles have a grade advantage of 1.0 percentage point, while those at higher quantiles increase their scores by 3.5 percentage points. Students that declared having reading habits showed a grade advantage compare to students that did not like reading. Quantile estimates show that this increase is higher for students at higher quantiles. In all four cases above, students at higher quantiles take greater advantage of a better study environment or quality of education.

Surprisingly, students that still live with their parents tend to have lower grades compared to the ones that live on their own. We suspect that there is a selectivity process in progress that is capture by this indicator variable. Students from out of town may have gone trough a tougher selectivity before coming here and thus are the best students from their home town. Moreover, those that do not live with their families have less of a social life and thus have less distractions and can dedicate more hours to study. Notice that above median students tend to have lower grades when living with parents compared to students at lower quantiles of the conditional score distribution.

Religion does play a role in students performance. As mentioned on previous section, those who declared having other creeds had a better performance compared to catholic students. Two quantile results deserve some attention here. Students who declared belonging to afro-religions not only had a grade advantage compared to catholics but also this advantage increases monotonically across quantiles of the conditional score distribution. Protestant students also have a grade advantage that vanishes as move along the conditional score distribution.

There is no consensus whether students employment improves or worsens student performance. Stinebrickner and Stinebrickner (2003) notice that individuals who fare well academically in school tend to be blessed with high levels of "motivation" that may also make them more likely than others to become involved with nonacademic activities such as work. Given that "motivation" is not fully observed, some of the variation in academic performance that should be attributed to differences in motivation may be mistakenly be attributed to differences in work status. Some studies like Turner(1994), Hood, Craig and Ferguson (1992) find that working a moderate number of hours is positively related to school performance. Paul (1982) finds that working is detrimental to academic performance in college, while Ehrenberg and Sherman (1987) find positive effects of working on-campus jobs but negative effects of working in off-campus jobs. Stinebrickner and Stinebrickner (2003) use OLS, fixed effects (FE) and instrumental variables (IV) estimates and find that OLS and FE estimators may tend to understate the negative effect that working has on academic performance. Both estimators find a positive effect of hours worked on grades. The IV estimates however indicates that working additional hours has a harmful effect on academic performance. Our estimates however indicate that working additional hours have a negative effect on students scores, on average and also for different quantiles of the score distribution. This result is stronger for students at higher quantiles compared to bellow median students.

As we saw before, students that attended public schools perform worse than private school students. The longer he/she stayed at the public school system the worse. Quantile estimates show that the negative impact of public school system increases as we move along the conditional score distribution, showing that students that have higher conditional scores suffer from a greater penalty for attending public schools compared to low conditional score students. As we explained before this result is a composition of worse socio-economic conditions of public school system, captured in our model, and also poor school quality, which is not captured in our analysis.

It is important to notice that the quantile coefficients for the students at the top 2% of the grade distribution are most of the time statistically insignificant. This is consistent with the findings in Levin (2001) and Birch and Miller (2006). Two conclusions may be taken from these results: first, there is a high degree of homogeneity among students who perform well at the test, second other factor such as motivation and study habits, not capture in our model, may account for these students performance. In both cases we conclude that scholastic achievement of Brazilian students in entrance tests, with grades in the ninetieth quantile could not be completely explained by independent variables included in the model.

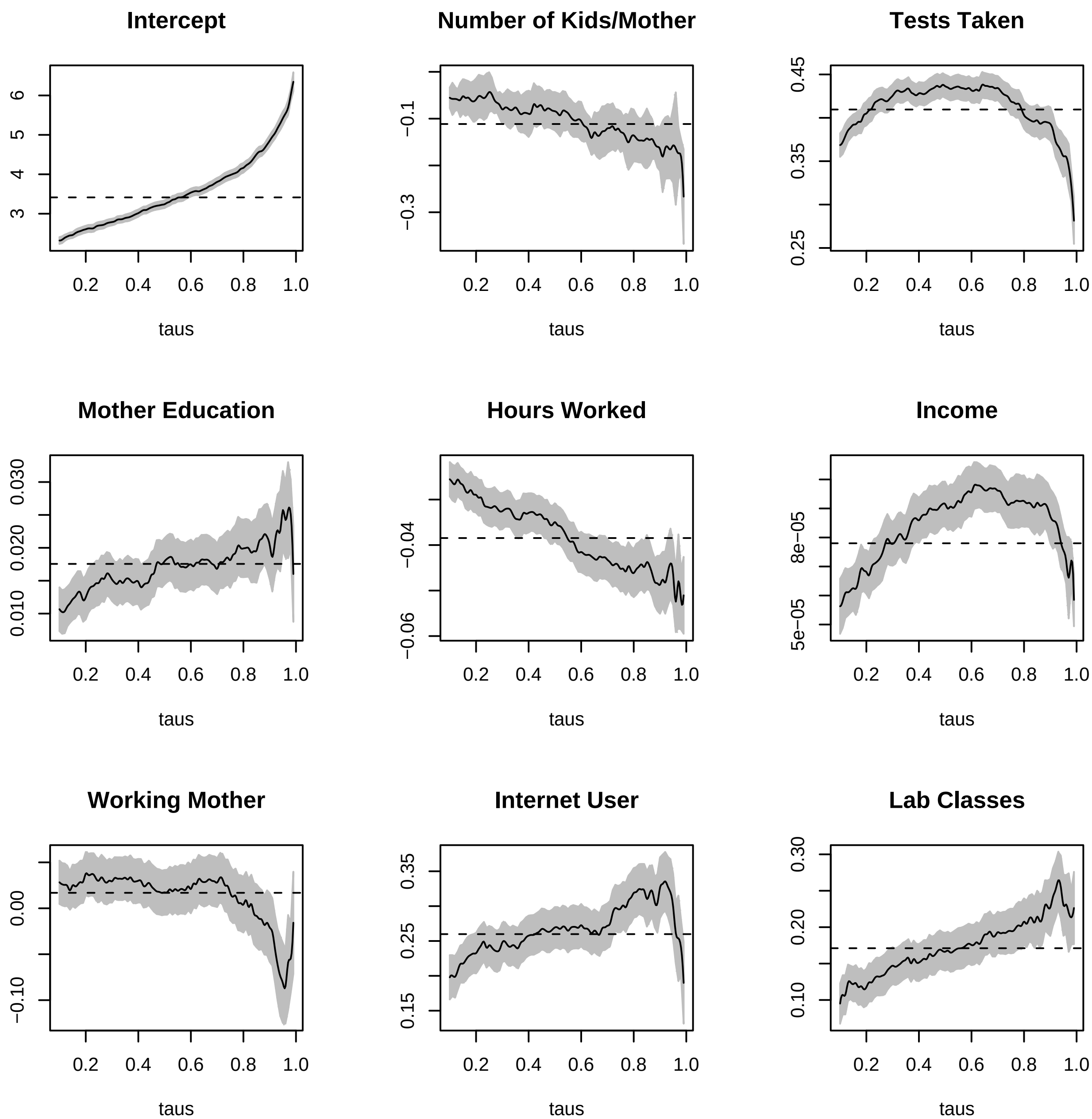

Figure 3: The above plots present the quantile regression estimates for each individual covariate indicated. The solid lines are the quantile estimates and the shade the 95% confidence intervals. The dotted line presents the Ordinary Least Square estimates.

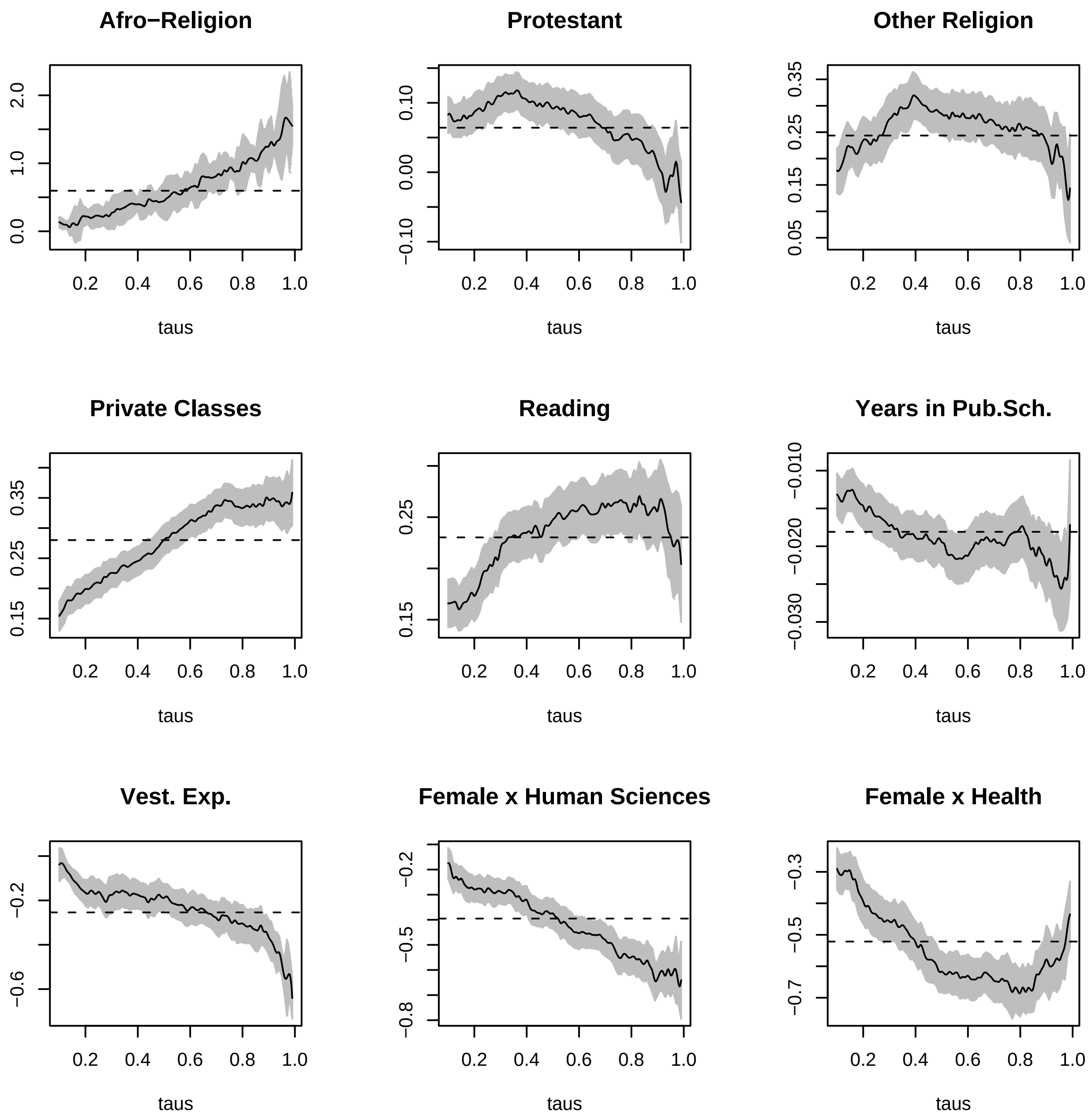

Figure 4: The above plots present the quantile regression estimates for each individual covariate indicated. The solid lines are the quantile estimates and the shade the 95% confidence intervals. The dotted line presents the Ordinary Least Square estimates.

5 Conclusions

In this paper we use an unique data set on students academic scores at Universidade Federal de Pernambuco (UFPE), which brings information on their standardized entrance tests scores, personal characteristics, such as age, gender, race, religion and family background, school attended and allow us to estimate the key determinants of their performance on the entrance test.

Using Least Squares and Quantile Regression we found some very interesting results. Female students perform worse than men on average. This result turns out to be stronger for women at the upper quantiles of the score distribution. Given that we did not find the same results on the empirical literature (see for instance McNabb, Pal and Sloane (2002)) we decided to control for the different University Departments. Women on average tend to perform better than men at predominantly male departments and on Medical School. The quantile results however show that on the Human Sciences and Social Sciences Departments women at upper quantiles suffer from a higher score gap compared to men. The presence of children lower women scores on average and across the conditional grade distribution.

Despite the fact that there is no consensus whether students employment improves or worsens student performance (Stinebrickner & Stinebrickner, 2003) we found that every additional hour worked decrease student scores by 9% on average. This penalty goes from 7% for the students at the lower quantiles of the conditional distribution to 12% for those at the upper tail.

Student that attended public school have their scores reduced by 4.7% for every additional year spent at the public system. This score gap is even worse for students at the upper quantiles of the conditional score distribution, who have their grades lowered by 6% for each additional year of public school. For those at lower quantiles, one more year at public school lower their grades by 3.8%. This result is closely related to the fact that public school students in Brazil have on average a worse family environment and poorer educational resources compared to private school children.

Family background and study environment are key determinants of student performance. Parents schooling impact positively on student scores. In fact, each additional year parents spend at school increases students scores by 4% on average. Moreover, students that have access to internet increase their scores by 60% while lab classes provide a 28% increment on their grades. Foreign language classes more than double their scores. Students at higher quantiles take greater advantage of extra educational resources such as access to internet, lab classes and private classes. These results are stronger for students at the upper quantiles of the conditional score distribution.

When it comes to personal characteristics our estimates show that married students have, on average, a 42% increment on their scores. Older students tend to have their scores reduced by 8%. The presence of kids reduce female scores 37% on average. Religion do play a role in students performance. Those who declared having other creeds had a better performance compared to Catholic students. Two quantile results deserve some attention here. Students who declared being afro-religion followers not only had a grade advantage compared to catholics but also this advantage increases monotonically across quantiles of the conditional score distribution. Protestant students also have a grade advantage that vanishes as move along the conditional score distribution. When it come to race only mulattos out performed white students having their scores increased by 10%.

Finally, our estimates showed that family background and educational resources are key determinants of students achievements. The effects however differ as we move along the conditional score distribution, showing that students react differently to same incentives probably due to factors not controlled in our model such as motivation and ability. Further research should work towards estimating this unobserved effects.

6 References

Arias, O., Kevin Hallock and Walter Sosa-Escudero (2001), "Individual Heterogeneity in the Returns to Schooling: Instrumental Variables Quantile Regression using twins data". Empirical Economics.20, 7-40.

Bassett, G. W., Tam, Mo-Yin, Knight, K. (2002). Quantile Models and estimators for data analysis. Metrika, 55, 17-26.

Birch, E. R., Miller, P. W. (2006). Student Outcomes at University in Australia: A Quantile Regression Approach. Australian Economic Papers, 45, n. 1, 1-17.

Clark, E. E., Ramsay, W. (1990). Problems of retention in tertiary education. Education Research and Perspectives, 17, 47-57.

Edie, E., Showalter, M. H. (1998). The effect of School quality on student performance: A quantile regression approach. Economics Letters, 58, 345-350.

Ehremberg, Ronald and Sherman, Daniel (1987). "Employment While in College Academic Achievement and Postcollege Outcome" Journal of Human Resources 22(1), 1-24.

Hallock, Kevin and Roger Koenker (2001) "Quantile Regression", The Journal of Economics Perspectives,15(4), 143-156.

Hood, Albert B., Craig, Andrew and Ferguson, Bruce (1992)"The Impact of Athletics, Part-Time Employment and Other Academic Activities on Academic Achievement" . Journal of College Student Development , 33, 447-453.

Koenker, R., and Bassett, G. (1978). "Regression Quantiles". Econometrica, 46 (1), 33- 50.

Koenker, Roger and Stephen Portnoy, (1997). "Quantile Regression", Office of Research Working Paper n.97-0100 College of Commerce and Business Administration, University of Illinois at Urbana-Champaign.

Kremer, M. and Levy, D. (2003), "Peer Effects and Alcohol Use among College Graduates", National Bureau of Economic Research Working Paperno.9873, National Bureau of Economic Research, Cambridge, United States.

Levin, J. (2001), "For Whom the Reductions Count: A Quantile Regression Analysis of Class Size and Per Effects on Scholastic Achievement, Empirical Economics, 26(1), 221-246.

Ma, L., Koenker, R. (2005). Quantile Regression Methods for Recursive Structural Equation Models. Journal of Econometrics, 134, 471-506.

Pin Ng and Jos Pinto (2003), Reducing High Attrition Rate in a Business Statistics Course Using an Interpretive approach Encompassing Diverse Teaching and Learning Styles, College of Business Administration, Nothern Arizona University Working Paper, Flagstaff,

Paul, Harvey (1982) "The Impact of Outside Employment Student Achievement in Macroeconomic Principles" Journal of Economic of Education , 13(summer), 51-56

Sampaio, B., Guimares, J. (2007). Diferenas de Eficincia entre Ensino Pblico e Privado no Brasil. Unpublished Paper.

Smith, J., Naylor, R. (2001). Determinants of Degree Performance in UK Universities: a Statistical Analysis of the 1993 Student Cohort. Oxford Bulletin of Economics and Statistics, 63, n.

1, 29-60.

Smith, J., Naylor, R. (2005). Schooling Effects on subsequent University Performance: evidence from the UK university population. Economics of Education Review, 24, 549-562.

Stinebrickner, R., Stinebrickner, T. R. (2003). Working During School and Academic Performance. Journal of Labor Economics, 21, n. 2, 473-491.

Tam, M., Basset, G. and Skhatme, U (2002) "New Selection Indices for University Admissions: A Quantile Regression Approach"in Y Dodge(ed), Statistical Data Analysis Based on the L1-Norm and Related Methods, New York, United States.

Turner, M. D. (1994). The Effects of Part-Time work on High School Students Academic Achievement. Paper presented at The Southern Economic Association Conference, Orlando, FL.

Win, R., Miller, P. (2005). The Effects of Individual and School Factors on University Students Academic Performance. The Australian Economic Review, 38, n.1, 1-18.