The London School of Economics and Political Science

From Preschool Provision to College Performances

Empirical evidences from a developing country

Paula Ines Giovagnoli

A thesis submitted to the Department of International Development of the London School of Economics for the degree of Doctor of Philosophy, London, June 2013

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I declare that my thesis consists of 29,265 words.

Statement of conjoint work

I confirm that Chapter 2 was jointly co-authored with Professor Sosa Escudero and Professor Alberto Porto and I contributed 60% of this work.

Statement of use of third party for editorial help (if applicable)

I can confirm that Chapter 1 and Chapter 3 were copy edited for conventions of language, spelling and grammar by Languague Centre, LSE and Chapter 2 was copy edited for conventions of language, spelling and grammar by Sue Redgrave.

Abstract

This thesis comprises of three stand-alone papers. The first paper exploits a natural experiment in Argentina to analyse what happens to maternal labour outcomes when there is an abrupt change in free public preschool provision. Using household survey data, the instrumental variable estimate shows that mothers work on average 9 hours more per week because her child is attending preschool. On the other hand, mothers for whom the access to public childcare was reduced did not change their labour supply, suggesting a shift in the mode of childcare from formal to informal.

The main goal of the second paper is to measure the effect of observable individual characteristics on the whole conditional distribution of performances. Quantile regression methods are shown to provide a flexible framework to model the interactions between observed and unobserved factors, which are the source of non-homogeneous effects on performance that alter its conditional distribution in subtle ways improperly summarised by mean OLS based methods. Using a database of students at public universities in Argentina, the empirical results strongly suggest the presence of heterogeneous effects, which leaves room to question whether relevant factors like parental education or secondary school type are stronger or weaker for certain individuals.

The third paper examines the role of labour market conditions on youth schooling behaviour using discrete time proportional hazards models. The findings show that, as predicted by human capital theory, labour demand has a significant effect on the hazard rate for dropping out of school. However, the results only hold for young males but not for females. The paper also tests whether each student's hazard rate for leaving school without completion changes autonomously over time. Using a non-parametric specification, the results indicate that the risk of dropout is increasing over time for both males and females.

Contents

List of Tables

List of Figures

Acknowledgements

I want to thank my supervisor, Dr. Jean Paul Faguet and my advisor, Dr. Jorn-Steffen Pischke, for their invaluable and untiring support. They not only provided intellectual guidance, but also offered moral support and encouragement during difficult times.

I thank the committee members for their very useful advice and in particular for comments and suggestions to improve the third chapter.

I am very grateful for the financial support from the Overseas Research Students Awards Scheme , DESTIN and a childcare grant from the Student Union. I want to thank participants at the "Work in Progress Seminar" for useful comments.

Special thanks to Giulia Ferrari, Julieta Trias, Borge Wietzke, Laura Jaitman and Fadi Hassan for helping me on personal issues.

I dedicate this thesis to my daughter, Lucia, who has been extremely patient during the last 3 years. I would not have completed this thesis without the support of my husband, German Capoulat. My parents, sisters and brother gave me invaluable help during the final year.

Introduction

The first chapter of this thesis is concerned with how the provision of preschool childcare affects maternal labour market outcomes.

The problem of non random selection into early education is overcome by exploiting two natural experiments that dramatically affected the probability of preschool attendance of children aged 4 in Cordoba - one of the three most important provinces in terms of population in Argentina - from 1996-99. In particular, I analyse the effect of two distinct and opposite policy changes: (a) the introduction of mandatory - free and public - kindergartens and (b) the closure of free and public kindergartens. These policies affected the preschool attendance of children differently depending on their month - year of birth and province of residence. Furthermore, given the way these new rules were implemented, there is no concern on their attendance being endogenously determined.

Using microdata from the EPH before and after the policies were implemented, I compare preschool attendance rates of children aged 4 in Cordoba with the rest of the country during times with and without the new rules in force. This difference-in-differences estimate showed that preschool attendance has increased by about 39% due to the mandatory policy. On the other hand, because of the closure of public kindergartens, there was a 46.5% decrease of the baseline attendance rate in Cordoba relative to the rest of the country.

In an attempt to control for specific shocks affecting Cordoba during these times, I also perform a triple difference-in-differences estimate. The results provide evidence that estimates are statistically similar to those obtained using just children aged 4.

Having shown that these new rules differentially affected enrolment into preschool, I exploit them to identify the causal effects of provision of preschool on maternal labour outcomes. Using the new rule as an instrument for actual attendance, I measured the impact among those assigned and complying with the new rule. The parameters estimated were interpreted as Local Average Treatment Effects (LATE).

The results indicated, first, that preschool attendance had an effect on the intensive margin; mothers work on average 9 hours more per week because a child was attending preschool. This new compulsory rule would affect those who had not been attending if the rule had not been set. That is, compilers in this case were those whose probability of attendance before the law was zero and now because of this new rule attend pre-school. They might come from more disadvantaged backgrounds and are likely to live in "less traditional" households. Second, mothers who have been induced to send their children to kindergarten are 18% more likely to work. However, this effect at the extensive margin was much less precisely estimated.

One concern with the identification strategy was that the implementation of policies happened very close to a financial crisis period. Cordoba could have been affected differently during and after the crisis compared to the rest of the provinces. In addition to a triple difference-in-difference strategy, and in an effort to create a counterfactual comparison group which better matched the treatment province in terms of pre-trend outcomes, I use the novel synthetic methods. Using synthetic weights, the main results still hold and suggest larger effects albeit less precise, which is likely because of the reduction in the sample sizes.

Unlike other studies, the importance of my results lies in the fact that by studying these two policies simultaneously, it is possible to learn that the effect of providing less or more preschool childcare on maternal labour outcomes is not symmetrical. That is, observing an increase in labour force participation, because of an increase in attendance, does not mean that a reduction in free public preschool would lead to a reduction in labour female participation.

The main goal of the second paper is to measure the effect of observable individual characteristics on the whole conditional distribution of performances. Quantile regression methods are shown to provide a flexible framework to model the interactions between observed and unobserved factors, which are the source of non-homogeneous effects on performance that alter its conditional distribution in subtle ways, which are improperly summarised by mean OLS based methods. This methodological framework is adopted and applied to the case of college students in Argentina, whose social and institutional characteristics, that combine free access, a flexible schedule and a diverse socio-economic composition of students, provide ample sampling variability, thus making it a relevant case study.

One of the main reasons for choosing this strategy is that in the case of educational policies, it is necessary to complement the standard educational production function approach, by studying not only the mean effects of observable variables but also their impact on the shape of the distribution of performances. This is relevant since educational policies are often expected to promote equality of opportunities and possibilities, and hence distributive outcomes matter. Also, if policy actions are oriented towards the less advantaged, or any other specific group, it is important to assess whether the impact of a policy measure is homogeneous for all students, or whether average effects are actually an imprecise summary of a more complex reality that may systematically benefit certain individuals more than others.

The empirical results of the paper indicate that, overall, effects are found to be less relevant in the top of the distribution, in the sense that all factors that contribute positively to performance (better family background, not having to work, etc.) are stronger in the bottom. Hence, policies that enhance the possibilities of students initially in the lower part of the distribution have the dual effect of increasing their absolute performances (through their positive effect) and reducing disparities due to their stronger effect in this group of students.

The third paper examines the role of the labour market conditions on youth schooling behaviour using a novel dataset from Argentina - Educacion y Empleo de los Jovenes -EJJ hereafter - that allows tracing of individual education histories from early in life in a retrospective manner.

The primary advantage of this dataset is that information on both the year of entrance at school and the year of exit from school are available, providing me with a unique opportunity to model 'dropping out' as a dynamic event in which the probability of leaving school can vary depending on the how many years a person has already spent in school. In addition, most school characteristics and socio-economic conditions of individuals at the time that schooling choices were made are known, providing us with a large set of controls.

The EJJ was implemented in 2005 to young people between 15 and 30 years old living in the main urban area of Argentina, Greater Buenos Aires. The fact that these cohorts were exposed to different labour market conditions by the time they reached secondary school is exploited in this study to identify changes in labour demand.

Using data from the EPH surveys, I construct unemployment rates as proxies for youth labour demand, by gender, level of education and year. These variables are matched to the EJJ sample to estimate the effect of youth unemployment on dropout by applying a discrete time proportional hazard model.

The findings show that, as predicted by human capital theory, labour demand has a significant effect on the hazard rate for dropping out of school. Young people who faced low unemployment rates by the legal working age were more likely to leave school than those who experienced worse conditions in the labour market. However, the results only hold for young males but not for females.

Parental education and age at entry showed to strongly affect the risk of leaving schooling not only for females but also for males. The study also tests whether each student's hazard rate for leaving school without completion does change autonomously over time. Using a non-parametric specification, the results indicate that the risk of dropout is increasing over time for both males and females.

1 Preschool and Maternal Labour Outcomes: Evidence from a natural experiment

1.1 Introduction

The question of whether mother's labour supply changes when there is an abrupt change in free public childcare provision has been extensively discussed in the literature. To estimate the effect of changes in public childcare provision on mother's labor supply, it is necessary to identify an exogenous change in childcare provision which is uncorrelated with the unobservable factors that jointly determine pre-school attendance and labour supply. This is because mothers who choose to work are very different from mothers who do not work in a number of important ways before they even enter the labour force.

Recent studies from the economic field attempt to account for this selection and simultaneity bias in a convincing way by identifying and exploiting "natural experiments" which provide some exogenous source of variation in provision of public (free) childcare (see Gelbach, 2002; Baker, Gruber and Milligan, 2008; Schlosser, 2007; Cascio, 2009; Fitzpatrick, 2010; Berlinski, Galiani and McEwan, 2011). Results for developed and developing countries are mixed. Furthermore, most of these studies have not moved beyond measuring the intention to treat estimates of public provision, thus do not provide a careful examination of how this effect might happen.

There are important distinctions between this study and previous works.

First, in this study, I do not only empirically examine the impact of public preschool provision on child attendance and maternal employment, but jointly investigate the actual effect of preschool attendance on maternal outcomes. Using a major policy change that dramatically affected the probability of preschool attendance of children aged 4 in an Argentinean province from 1996 to 1999 , I am able to uncover the causal effect of interest. In particular, I am able to estimate the impact of the new policy on preschool attendance among those affected by the policy change. The idea is that after conditioning for a set of exogenous variables the change in the probability of preschool attendance induced by new policy is uncorrelated with the error of the maternal labour outcome equation.

Second, unlike other studies, I was able to investigate changes in preschool attendance in different directions. In addition to studying how increases in attendance affect the maternal labour market, I also analyze whether any reduction in preschool availability impacts on maternal labour outcomes. Results from the former reinforced findings obtained by Berlinski et al. (2011). These authors used a regression discontinuity design exploiting the traditional cut-off rule. In order to validate their analysis, they had to show that parents did not choose the season of birth for educational purposes. Official records from birth registrations are used to test this assumption. However, due to the availablity of this data they were only able to analyze a period different from the one used for the estimations and they were only able to include a sub-set of provinces, leaving Buenos Aires out of the analysis.

Third, I allow a much richer choice set for the labour outcomes providing a better understanding of how the policy could affect the maternal labour supply.

Fourth, by using synthetic methods, I was able to choose a better control group of provinces in terms of pre-outcome trends.

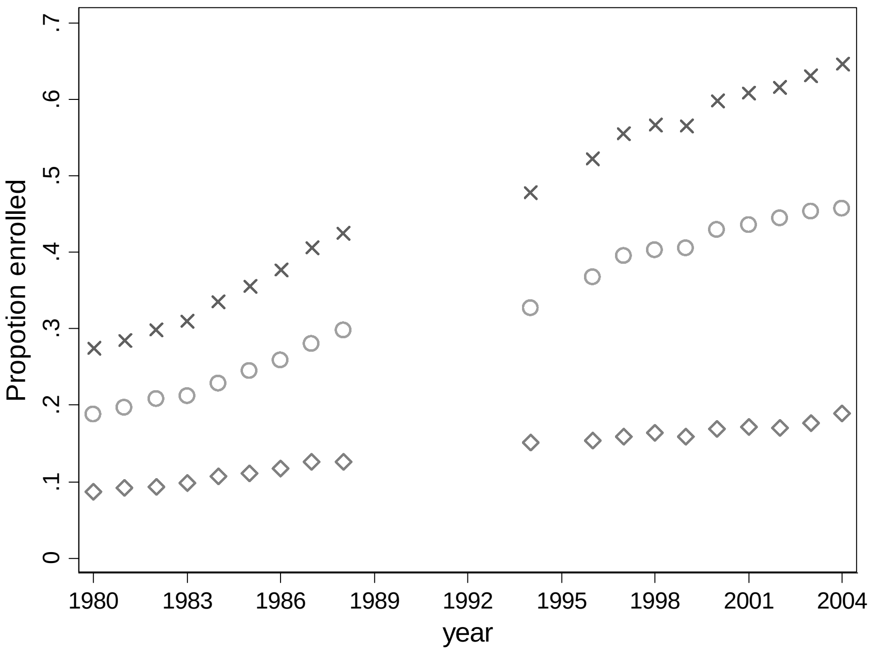
The remainder of the paper is organized as follows. Section 1.2 describes briefly preschool education in Argentina and uses different sources to analyze preschool enrolment. The implementation of the new policy in Cordoba is explained followed by a discussion on how it could be related to the labour supply of mothers. Section 1.3 presents the survey data used to estimate the parameters of interest. The section starts by performing a description of the policy change and the labour market conditions for the period surrounding the new policy's implementation. Section 1.4 shows the empirical results including some robustness checks while section 1.5 concludes.

1.2 Institutional Background and the Policy Change

1.2.1 Preschool education in Argentina

Public preschool education in Argentina is composed of 3 levels - known as Salas - for children aged 3 to 5. In order to be enrolled in a given Sala the child should meet the minimum age requirement by June 30th (the cut-off date) of the school year. This rule is the same across the country meaning that to be enrolled in Sala 5 (4, 3) the child should turn 5 (4, 3) on or before June 30th of the school year. The school year (i.e. attendance period) starts in March and ends in December uniformly across provinces. The enrolment period for a given school year takes place by the end of the previous year. While enrolment in Sala 5 has been compulsory since 1993, Sala 4 and 3 are optional.

FIGURE 1.1
Proportion of children aged 3-5 in Preschool Education. Total Country



References: x Total Attendance, O: Public Attendance, <> Private attendance. No data available at national level for 1995.

Sources: Enrolment data 1980 -1988: MEJ SE DINIDETE Centro Nacional de Estadísticas de la Educacion. División Sistematización de Datos. 1994 Schools and Teachers National Census. Direccion General Red Federal de Informacion Educativa 1996-2004 Relevamientos Anuales. Direccion General Red Federal de Informacion Educativa Population Data: Census 1991 and 2001.

I draw from a variety of data sources to present an historical analysis of preschool attendance data[1]. I supplement administrative data with micro census data for 1991 and 2001.

The analysis of this data suggests a rapid growth trend, similar to that observed in developed countries (see, for instance, Havnes T. and Magne Mogstad, 2011).

[1]The series used in this paper were not ready available. I want to thank the National Ministry of Education and Santa Fe and Cordoba provincial Ministries of Education for providing me with the time series.

Figure 1.1 displays gross enrolment rates for children aged 3-5 over the period 1980-2004 in Argentina separately by total, private and public enrolment. Total coverage was around 25 percent in 1980 and rapidly rose to 65 percent by 2004.[2]

In terms of the number of children attending preschool this expansion means that the system as a whole enrolled 810 thousand more children in 2004 than it did in 1980.

The expansion is mainly driven by enrolment at public institutions. The latter operate 5 working days a week, they are free of charge and provide three and a half hours of childcare per day. Private institutions set a fee and offer between 3 to 8 hours of childcare depending on parental preferences.

The national rising trend in attendance can be observed across all provinces. In Cordoba Province, however, this trend was interrupted between 1996 and 1999 (as shown in Figure 1.2). During that period, a new policy on preschool enrolment had been introduced by the Provincial Government.

Before turning to the next section in which I describe the policy in detail, it is worth mentioning that Cordoba is the third province in terms of total population, after Buenos Aires and Santa Fe and the second one in terms of total children according to the Official Population Census carried out in 2001. Out of the total of children aged less than 6, around 35 percent live in Buenos Aires, 8 percent in Cordoba, and 7.5 percent in Santa Fe. More importantly, by 1996 Cordoba was one of the leading provinces in terms of preschool developments with high public coverage of at least 80 percent of the total supply.

1.2.2 The Policy Change in Cordoba

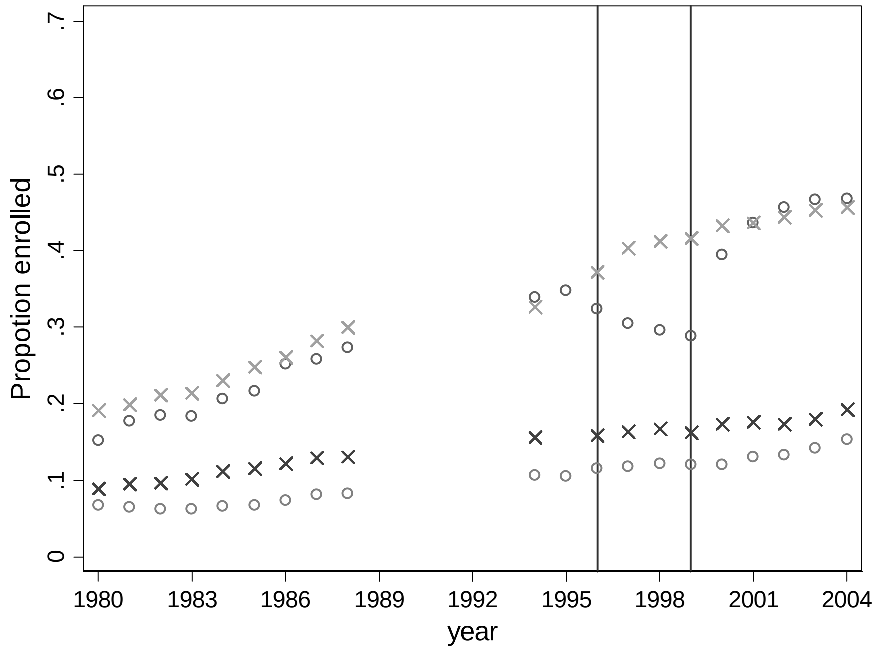
In 1996, a new preschool admission policy was suddenly established in Cordoba by the Provincial Government, allowing only children aged 5 to attend public preschool education. While public Salas 5 were kept open, Salas 4 were all closed. In addition, the province also changed the cut-off date to start Sala 5, extending it to August 31st of the school year[3]

[2] I would like to graph these trends separately by Salas 3, 4 and 5. Unfortunately, this distinction is only available from 1996 and for a select group of provinces

[3] Although the rule implicitly imposed the closure of Salas 3 as well, these were not common at that time, for which reason the policy mostly affected those families with children aged 4. Furthermore, the policy was popularly known as "Cierre de Salitas de 4". In fact, Sala 3 enrolment in Cordoba and ROC was extremely low during the 1990's - below 3 percent when excluding Greater Buenos Aires and the Capital (see Appendix Table 1.7).

FIGURE 1.2

Proportion of children aged 3-5 in Preschool. Cordoba versus Rest of Country.



References: O: Cordoba, X: Rest Of Country. Rest of Country. Upper lines refer to Public Enrolment. Lower lines refer to Private Enrolment.

Source: As in Figure 1 except for enrolment data for Cordoba 1995: Mensaje Apertura Sesiones Legislativas (cited by Abratte, 2007).

These new rules induced two main changes for some children of preschool age living in Cordoba if compared to similar children living in the rest of the country or in Cordoba during previous years:

1. The extension of the cut-off date means that Sala 5 will also enrol children who turn 5 between July 1st and August 30th of a given school year. Preschool education will become compulsory for them. Without these rules in force, however, these children represent a subgroup of those who are eligible for Sala 4 with voluntary enrolment. They would have been the oldest children in the cohort of Sala 4. Throughout the paper, I will refer to these children as “Group 1”. Note that this rule affects the 1996 enrolment period onwards, therefore compulsory education is effective for the group that starts in

1997.

2. The closure of Salas means that there is no more public preschool available for children younger than 5. I will refer to these children as "Group 2". Note that this rule affects the 1996 attendance period, therefore affecting the group of children who were already enrolled and expected to start Salas in the 1996 school year.

In 2000 these rules were abrogated and the old one was reinstated. This implied the re-opening of Salas as well as the re-adjustment of the cut-off date to June 30th. [4].

A possible concern is that the Government decided to pass the new rule in response to low preschool enrolment, invalidating the possibility to exploit these new rules as an exogenous determinant of attendance unrelated to parental preferences. For instance, if in the year prior to the implementation an economic crisis affects labour market conditions and induces parents to take out their children from school, observing a decrease in attendance does not necessarily mean parent's response to Government new rules but instead the Government's response to parent's behaviour.

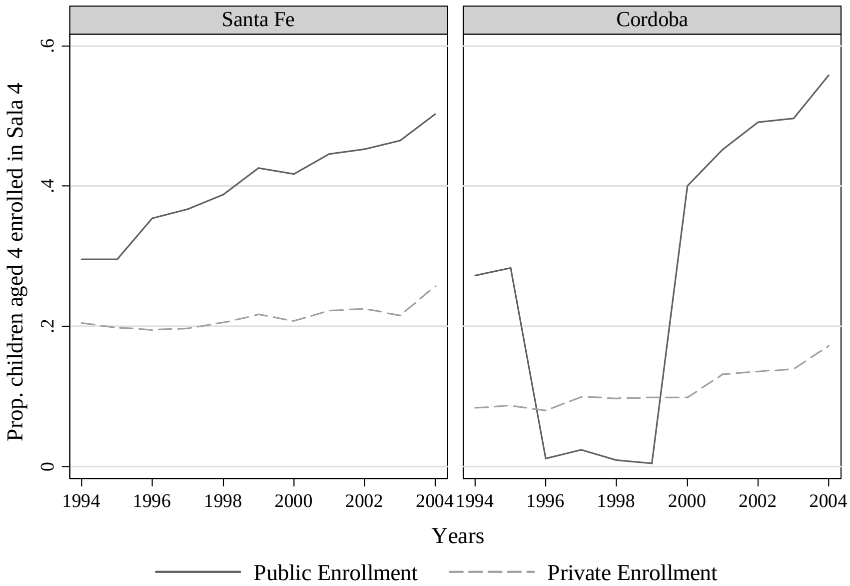
Figure 1.3 provides evidence that the situation is unlikely. The figure compares the proportion of children enrolled in Sala 4 at public and private institutions in Cordoba (right panel) and in Santa Fe (left panel). It is reassuring to observe there is no lower enrolment just before the implementation of these new rules. In fact, in 1995 around 30percent of 4-year-olds were enrolled in Sala 4 - more than 16,000 children - in Cordoba. A year later, this figure dropped close to zero - 637 children - and was kept virtually nil until Sala 4 was re-opened in 2000 incorporating 22,000 children in preschool education. Table 1.8 in Appendix reports total number of children enrolled in Sala 4 by year and by province. On the left panel, Santa Fe province data indicates rising trends in public Sala 4 during the 1990's.

The process of how these rules were implemented and the reasons provided in the Government documentation also contribute to ruling out any possibility that the policy change was implemented as a response to parental behaviour. In fact, financial budget restrictions have been stated as the main reason for closing Salas by the Provincial Government. The announcement was made just a couple of weeks before the school year started during the annual teacher's

[4]See Appendix. Section 1.6.1 for a description of Decrees.

FIGURE 1.3

Proportion of children enrolled in Sala 4. Santa Fe versus Cordoba.



Sources: 1994: Teachers and School Census 1994. 1995: Mensaje Apertura Sesiones Legislativas (cited by Abratte, 2007) 1996-2004: Dirección General de Información y Evaluación Educativa.

general meeting on February 26th 1996. It was issued in a circular by the Provincial Ministry of Education. Most preschool teachers lost their jobs and others were reallocated depending on their category - supply teachers or tenure, respectively. Regarding the reasons for making preschool mandatory for a subgroup of children, the Provincial Government refers to enhancing children's development and knowledge early in life. In fact, the new Decree 1628/96 issued to modify Art. 1 of Decree 8912/88 in order to regulate enrolment rules states as follows:

> "Considering that research and ethnographic analysis related to preschool insertion into educational institutions provided evidence of important improvement in terms of child behaviour, such as more and better use of cognitive strategies applied to solve problems; socialization processes with better ability to adapt to norms; better development of autonomy and earlier needs of communication (through reading and writing) than expected....And given the coincidence of pedagogical innovations the province is carrying out, ...the Government decrees that ....Art 1: ...[..].. "In preschool institutions, only children aged 5 before August 31st of the school year will be admitted[5]".

This decree recognizes preschool education as a developmental tool imposing compulsory attendance. However, the provision of free preschool for all children aged 4 was not possible at that time due to financial restrictions. The Provincial Government gave access only to a subgroup of children aged 4 (those whose turn 5 during July and August of the school year).

The changes in these enrolment rules create a presumably exogenous variation in preschool attendance that can be used to estimate its effect on the maternal labour market provided that the potential affected groups can be distinguished by the potential control group in the data. .

The policy change and maternal labour supply

For those mothers planning to send their kids to public preschool the unexpected closure of Sala 4 would make participation in the labour market less attractive as their opportunity cost

5

The same decree also introduces new rules for children who enrol in first grade of primary school. Again, the cut-off date is extended accepting children turning 6 before August 31st (instead of June 30th). This meant that in 1997 - during the first year of implementation - the size of the first grade was on average 16 percent bigger than previous cohorts. Unfortunately, there is no data available to exploit this change in classroom size of working increases. Now, these mothers would work only if the market wage is equal or higher than the cost of paying for taking care of their children. That is, because the reservation wage increased, mothers would substitute work (in the market) for work at home (taking care of their children who now cannot attend Sala 4). However, closure of Sala 4 not only produces a substitution effect, but could also produce a negative income effect. This is because mothers who, in the absence of the policy, would have used public Sala 4 instead of paying for childcare, and would perceive a reduction in their income making participation in the labour market more likely. Under a static model of labour supply, any of the effects could dominant, depending on the type of mothers affected. Furthermore, it could also be the case that the affected mothers develop informal childcare arrangements in order to prevent any effect on their own labour supply.

On the other hand, the reverse policy - making attendance to preschool mandatory - would act in the opposite direction, with a substitution effect raising hours of work for those already working and a positive income effect reducing hours of work if "time at home with children" is a normal good.

1.3 The Data

This section describes the official Urban Households Survey -*Encuesta Permanente de Hogares* (EPH) - from which the potential affected groups can be distinguished by the potential control group in the data before, during and after the changes in the rules, making it possible to exploit this “natural experiment” (Meyer, 1995).

The EPH has been regularly conducted twice a year during May and October from 1983 until 2003 by the National Institute of Statistics and Census (INDEC)[6]. The sample size was set to achieve - with 95percent confidence - an error of 1percent in the unemployment rate within each urban conglomerate (Ravallion M, et al., 2001). In each round about 80,000 to 100,000 individuals were interviewed providing data on the labour market situation, education, income and socio-demographic variables as well as data on household characteristics and composition. The survey has incorporated new areas over time to widen its coverage. By the late 1980’s,

[6]Since 2003, INDEC changed the methodology of the survey and questionnaires. It is now conducted 4 times a year under the name of Encuesta Continua de Hogares.

33 urban areas with more than 100,00 inhabitants were included, representing 71 percent of the Argentine urban population. The availability of this data provides the opportunity to assess trends on main labour market outcomes comparing Cordoba province with the rest of the country over a long period (see Section 1.3.2).

For the main analysis, however, I am only able to exploit EPH survey data from 1995 to 2001. The choice of the period is motivated by two considerations. First, only since 1995 did the survey collect information on the exact date of birth for each individual in the sample. This information is vital to identify children potentially affected by the new rules. Second, information on preschool attendance was not reliable before this period. Actually, most agglomerates reported missing values in this variable for children aged 3-5 before 1994.

To retain the maximum number of observations as possible, I pooled repeated cross-sections of individual level data from both waves, May and October[7]. If households (mothers/children) appear in more than one wave their recurrence is treated as coincidental and ignored. The usual assumption implicit on the analysis of this type of data is that the unobservable individuals effects are drawn from the same population distribution across periods to avoid composition bias. (See Wooldridge, 2013 : 129 and Cameron and Trivedi, 2013 : 47-75)

1.3.1 Descriptive evidence of the policy change

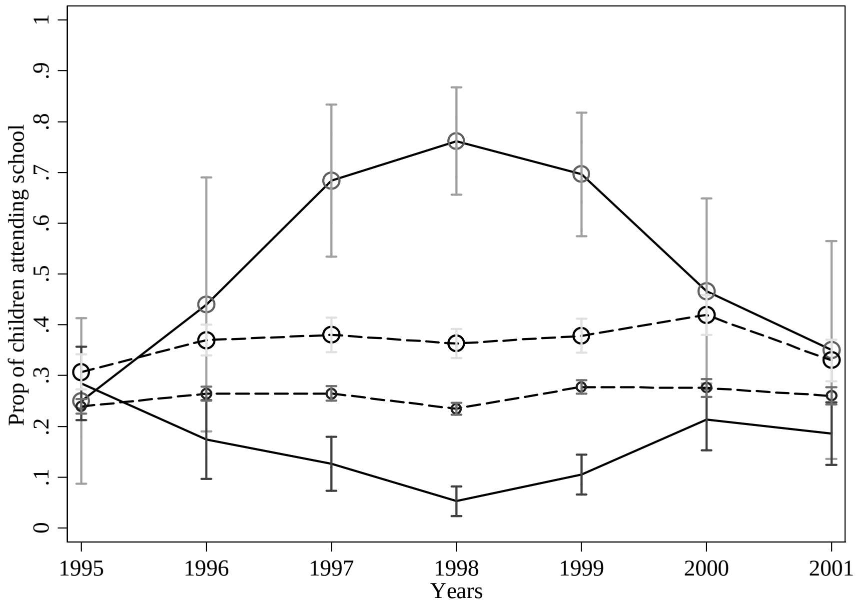
Figure 1.4 plots the percentage of children aged 4 attending school by year and the 95 percent confidence intervals using micro data from the urban household surveys. The sample consists of all children between the age of 48 to 59 months by June 30th of each survey year with no missing value on attendance.

These estimates are presented for Cordoba (the "treated" province, solid lines) and the rest of the country (ROC, "control" provinces, dash lines). I further divide the sample according to children's age - those aged 58 and 59 months (Group 1, oldest children in their cohorts) versus those aged 48 to 57 months (Group 2, youngest children in their cohort).

This data reflects the sets of new rules clearly. Attendance rates in Cordoba and ROC were similar prior to the implementation, around 25-30 percent. Between 1997 and 1999 these

[7] In 1997 and 1998, the survey was also conducted during August. The analysis performed in section 1.4 includes this data and control for year-wave effects. Results excluding these waves are similar, although less precise.

FIGURE 1.4
Proportion of children aged 4 enrolled in preschool. Cordoba versus Rest of the Country.

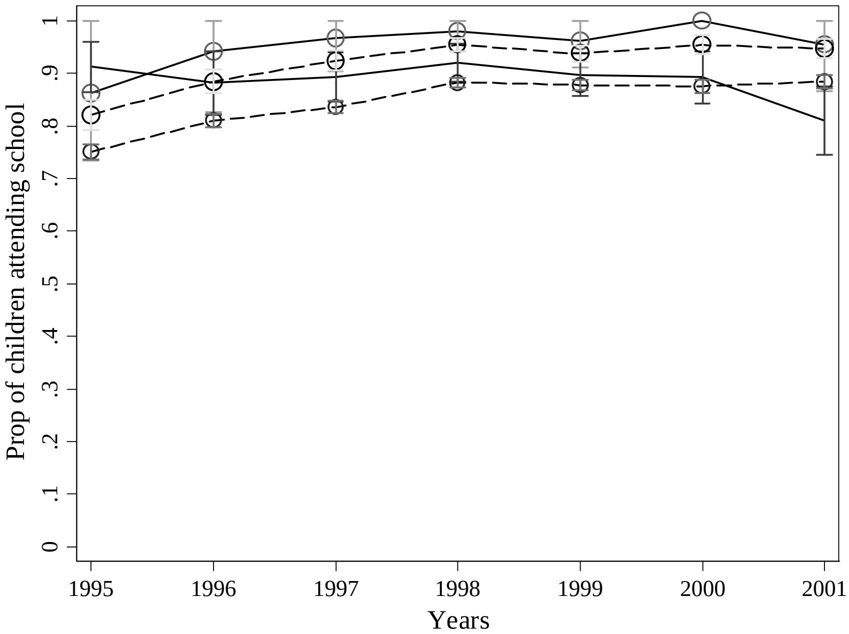


Source: EPH. 1995-2001. May and October waves. Sample of children aged 4.

Note: "O" refers to Group 1 children (born between July 1st and August 31st). Solid lines refer to Cordoba. Dashed lines refer to ROC. Markers refer to 95 percent confidence intervals.

FIGURE 1.5

Proportion of children aged 5 enrolled in preschool. Cordoba versus Rest of the Country



Source: EPH 1995-2001. May and October waves. Sample of children aged 5 (that is, 60-71 month of age).

Note: "O" refers to Group 1 children (born between July 1st and August 31st turning 6). Solid lines refer to Cordoba. Dashed lines refer to ROC. Markers refer to 95 percent confidence intervals.

rates increased by almost 40 percentage points among Group 1 children while they decreased towards zero for Group 2 children in Cordoba[8].

The descriptive analysis reveals some interesting facts. Compliance with the new rules is imperfect in both cases. That is, some children assigned to mandatory attendance, did not attend. Conversely, some children affected by the closure of Sala 4 are still attended preschool anyway, probably at private institutions. Unfortunately, this data does not allow me to identify if the child attended public or private preschool education. It can be argued that since administrative data shows no more than 10 percent of children were attending private institutions during the pre-intervention period, this provides some evidence that trends observed in the survey data are mainly driven by public attendance. Furthermore, Figure 1.3 reveals that no significant changes in private attendance for children aged 4 occurred during the implementation of new rules, suggesting substitutions between public and private enrolment, if any, have been marginal. Notwithstanding, even if administrative data did not show any increases at all between 1996-99, and if it prevented a decline in enrolment, it suggests that private attendance could actually have been lower if the new rule had not been passed. If that is the case, the new law could actually have lead to substitutions between public and private enrolment.

Finally, Sala 5 was kept open in Cordoba during that time. As data shows in Table 1.5, preschool enrolment for children aged 5 is fairly stable during the whole period. There are no significant differences across groups and years. The data also shows that despite attendance being compulsory for everybody at this age, non compliance is an issue as well but less severe.

1.3.2 The labour market for the period surrounding the policy change

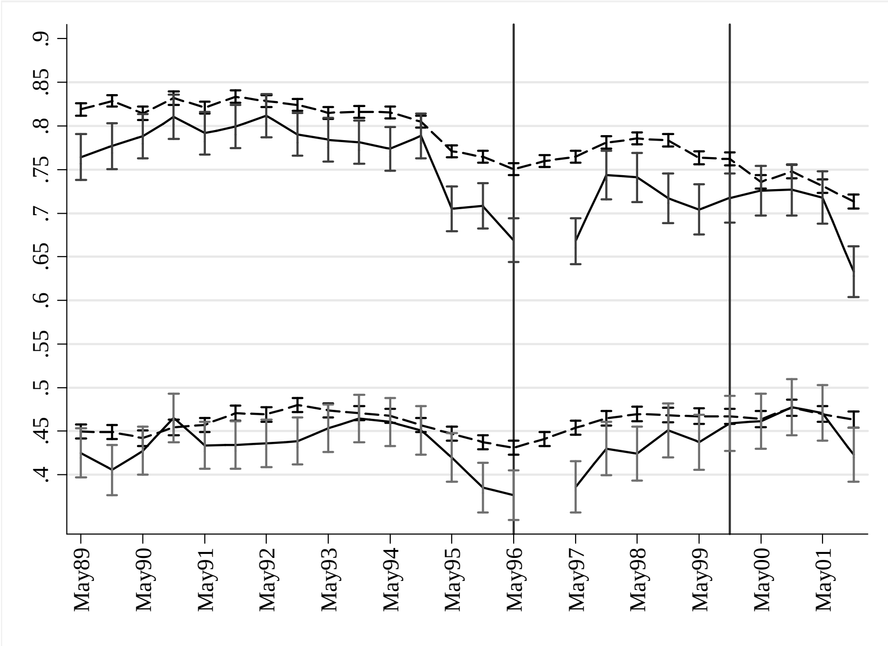
This sub-section briefly analyzes the general labour market situation around the time of the implementation of new rules using the official household survey data . The aim is to compare Cordoba with the rest of the country from 1989 to 2001.

The analysis includes all men and women between aged 18 and 59 (49 in the case of women) above high school completion age and below the standard retirement age. The sample

[8] There was no survey carried out in Cordoba Capital during October 1996. Estimates shown for Cordoba 1996 are calculated using the May wave only for the Capital and Rio Cuarto agglomerate. Standard errors are wider than the rest of the years. In fact, the proportion of children from Group 1 attending school in 1996 in Cordoba is very imprecise, from 0.19 to 0.69. In 2001, a sample reduction in most of EPH agglomerates by almost half produced less precise estimates as well.

FIGURE 1.6

Employment Rates. Men and Women aged 18-59 (49 for women).



Source: EPH. 1989-2001. May and October waves.

Note: Upper lines refer to men and lower lines refer to women. Solid lines refer to Cordoba and dashed lines refer to Rest of the Country. Vertical lines indicate the period of the new rules were applied.

excludes those observations with missing values on employment status. Due to data restrictions these trends cannot be shown by subgroup of mothers with children of different month/year of birth as the exact date of birth is not available for most of the pre-treatment period (i.e. no date of birth from 1989 until 1994).

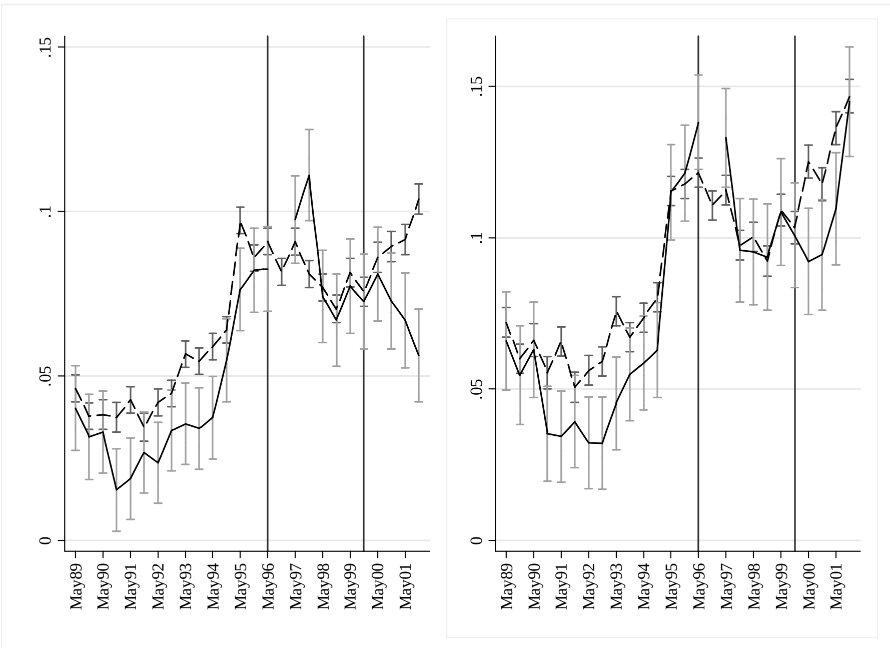
Figures 1.6 and 1.7 plot employment and unemployment rates in Cordoba and ROC for men (upper lines) and women (lower lines) with their corresponding confidence intervals. The two long vertical reference lines in May 1996 and October 1999 indicate starting and ending new rule dates.

The first thing to notice is that while male employment is almost double female employment, the latter presents levels similar to those in developed countries.

In December 1994 theTequila Crisis - originated in Mexico but lead to significant spillover effects to Argentina during 1995 - produced a collapse in financial flows and investment leading

FIGURE 1.7

Unemployment Rates. Men and Women 18-59 (49for women). Cordoba and Rest of the Country.



Source: EPH. 1989-2001. May and October waves.

Note: Left panel refers to unemployment rates for women. Right panel refers to unemployment rates for men. Solid lines refer to Cordoba and dashed lines refer to Rest of the Country.

to a deep recession. This crisis stands out clearly in the data. Since May 1995, employment rates show downward trends for both, men and women. It is worth noting how the decline trend in Cordoba is significantly deeper compared to the ROC while showing a faster recovery allowing the province to converge to similar employment rates than the ROC by 2000. This suggests that Cordoba experienced stronger shocks during the period under analysis.

Furthermore, while at the beginning of the 1990s Cordoba had registered lower unemployment rates than the ROC, during the crisis period a rapid increase made the province reach similar rates to the ROC, exceeding two digits for the case of men[9].

This analysis is suggestive of the need to take into account that the period during which new rules on attendance were implemented in Cordoba overlaps most of the crisis and post recovery period. A natural concern is how to properly isolate the effect of new rules from the effect of stronger shocks experienced in Cordoba itself. Using the ROC labour market behaviour as a sole contrafactual situation for the Cordoba labour market might not be enough to properly identify the effects of the new rules on maternal labour outcomes.

1.4 Empirical Strategy and Results

1.4.1 The effect of new rules on preschool attendance

As a starting point to assess and measure the effect of the reform on preschool attendance I estimate difference-in-differences models comparing preschool attendance rates of children aged 4 in Cordoba and the rest of the country during times with and without the new rules in force. The generic estimating equation (1) is:

$$A_{ipt} = \beta_0 + \beta_1 Compulsory + \beta_2 Closure + \alpha_p + \gamma_t + \lambda_1 Age_{Gr} + X_{ipt} + \varepsilon_{ipt}$$

where i denotes children, p provinces (urban agglomerates), t time. The dependant variable, A_{ipt}, is equal to 1 if the child attends preschool and zero otherwise; *Compulsory*is a dummy for being affected by the new rule of compulsory preschool, that is, she/he is a 4 year old (Group 1) child residing in Cordoba from 1997-1999, *Closure* is a dummy for being affected by the new rule of preschool closure, she/he is a 4 year old (Group 2) child residing in Cordoba from 1996 -1999; α_p denotes agglomerate fixed effects[10] to account for permanent

[9]Note, however, that in this case bigger standard errors of these estimates made it impossible to clearly identify significant differences in unemployment rates during the treatment period.

[10]In some provinces the survey includes more than one agglomerate. The incorporation of these agglomerates

differences in attendance rates across agglomerates, γ_t captures time fixed effects (a dummy for each survey round), Age_{Gr} is a dummy variable indicating the age group the child belongs to, X_{ipt} is a vector of control variables that includes: a dummy equal to 1 if the child is a boy, controls for household characteristics: a set of variables denoting the total number of siblings in the following age groups: 0-2, 3-5, 6-13 and 14-18, controls for mother's characteristics: mother's education dummies, mother's age dummies, and a dummy equal to 1 if the spouse is present. ε_{ipt} is the error term. The coefficients of interest are β_1 and β_2 which captures the effect of these new rules on preschool attendance. Standard errors are clustered by year agglomerate. (Bertrand, Duflo, and Mullainathan 2004).

Before shifting the attention to the empirical results, it is worth noting that a source of potential bias could arise if some families with children of preschool age living in Cordoba decided to migrate to other provinces in order to get public preschool education. This migration would invalidate any comparison of groups over the years. Within the Rubin Causal Model (Rubin, 1974), this means the new rule will not be independent of the potential treatment assignment, violating the Stable Unit Treatment Value Assumption (SUTVA)[11]. Although there was a high degree of uncertainty about how long the new rule would prevail - making this kind of behaviour less likely, with the data at hand I am not able to empirically test this hypothesis. Despite closure of Salas being an unexpected event in 1996, it might happen that in the years following the new rule implementation, parents react to the rule beforehand. If those who migrate are not a random subset of potentially affected families but instead are families with higher preferences towards education within a skilled labour force, the composition of eligible (and non-eligible) groups would change. I can argue, however, that the EPH sample coverage of Cordoba Capital and Rio Cuarto towns, both are in the middle of the province - north and south, respectively - and presumably it could be more difficult for parents to send their children to a preschool institution located in a neighbouring province. Therefore, throughout the paper I assume that SUTVA holds.

Regression Results

---

into the survey differs across time within the same province, leading to an unbalanced panel of urban agglomerates. For instance, Mar del Plata in the Province of Buenos Aires, Concordia in the Province of Entre Rios and Rio Cuarto in the Province of Cordoba were incorporated into the survey in October 1995.

[11] See also Holland (1986) for an elaborated exposition of the Rubin Model for Causal Inference.

Table 1.1 presents the main results for the sample of children with no missing values for preschool attendance and the exact date of birth variables.

While the first row shows the coefficient estimates of the effect of the new compulsory enrolment rule on preschool attendance, the second row shows the effect of closure of preschool. Results from the difference-in-differences and triple difference? models are reported in columns (1) to (3) and columns (4) to (6), respectively.

Clearly, the introduction of the new rules had sizable effects on preschool attendance behaviour.

Looking at the difference-in-differences specification, there is a highly statistically significant estimate of the new rule of mandatory preschool for children aged 4 (Group 1). Attendance rates have increased by about 39 percentage points due to the new rule. Furthermore, results from the second row show that the probability that a child aged 4 (Group 2) attends preschool decreased by 12.3 percentage points in Cordoba relative to the rest of the country, representing a 46.5 percent decrease of the baseline attendance rate. These estimates from column (1) are based on a sample of "all children" aged 4.

In column (2), I present similar estimates but now restricting the sample to children living in households where the mother can be identified. I consider this subsample because it is this subgroup of data which I can assess for the effect of preschool attendance on “maternal” employment. The mother can be identified using the question of the survey on the relationship between the head of the household and the household members. Therefore, the restricted sample is composed of all children who reported being the “son/daughter” of the head of household from households where both parents are present or the mother is the head[12] . Results are very similar to those shown in Column (1). Nevertheless, the effect of closing Sala 4 is slightly higher while the effect of the mandatory rule is lower compared to estimates using the “all children” sample. Closing Sala 4 would affect those who had been attending under the old rules but because of the new rule they cannot attend. Conversely, the new compulsory rule would affect those who had not been attending had the rule not been set. That is, compilers in this case are those whose probability of attendance before the law was zero and now because

[12]Note that the sample of “all children” in column (1) includes also children living in “less traditional households”, most of them grandchildren of the head of the household

of this new rule attend preschool. As discussed in Berlinski, Galiani and McEwan (2011) they might come from more disadvantage backgrounds and are likely to live in "less traditional" households.

One may argue that there are other factors not properly controlled for explaining these findings. It might not be the new rule per se that induces changes in the proportion of children attending preschool but instead other socioeconomic characteristics or specific shocks affecting Cordoba during that time.

In an attempt to explore the existence of omitted factors, in column (3) I re-run the regressions for the restricted sample but now include additional observables characteristics such as: mother's education dummies, mother's age dummies, a dummy for mother is single and total number of sisters/brothers between 0-2, 3-5, 6-13 and 14-18 living in the household. I found that estimates change little while slightly gaining precision. In columns (4) to (6) I also include an additional comparison group - children aged 3 or 5 from households not affected by the new rules. This leads to the triple difference-in-differences estimate. As explained earlier, the difference in attendance rates for children aged 4 between Cordoba and the rest of the country and before and after the implementation of the new rules is contrasted with the equivalent difference among children aged other than 4[13]. Column (4) shows that estimates are statistically similar to those obtained using just children aged 4 (the DD estimates). This is consistent with Figure 1.5 presented in the previous section in which attendance rates for children aged 5 in Cordoba did not reflect any difference across years in comparison to attendance rates of children of the same age living in the rest of the country. The results of DDD estimates also hold in the sample of children in which the mother is identified and when I added children's and mother's controls. See coefficients estimated in column (5) and (6).

In sum, results from this section provide strong evidence that the new rules play an important role in explaining changes in preschool attendance in Cordoba for children aged 4 during the 1996-1999 period. The next goal is to evaluate whether these rules have an effect on the labour supply of mothers.

[13]In households where there is a child aged 3 and a child aged 5, I consider the youngest child for the analysis. Children who turned 5 in July or August the previous year were excluded from the analysis as in Cordoba this group had to attend first grade. As most of the children already attend school if the new rule produced any difference, it was difficult to capture this with the data at hand.

TABLE 1.1

Estimates of the Effect of New Rules on Preschool Attendance

Variables	(1)	(2)	(3)	(4)	(5)	(6)
Compulsory Attendance	0.389***	0.365***	0.356***	0.337***	0.274***	0.251***
	(0.0435)	(0.0520)	(0.0439)	(0.0588)	(0.0832)	(0.0804)
Preschool Closure	-0.123***	-0.144***	-0.145***	-0.132***	-0.150***	-0.148***
	(0.0309)	(0.0336)	(0.0317)	(0.0396)	(0.0377)	(0.0367)
Observations	32,061	25,323	25,323	95,770	62,283	62,283
R-squared	0.230	0.226	0.249	0.522	0.479	0.490
Restricted Sample	No	Yes	Yes	No	Yes	Yes
Additional controls	No	No	Yes	No	No	Yes
Sample Child 3-5	No	No	No	Yes	Yes	Yes

Source: Encuesta Permanente de Hogares, May 1995 to October 2001.

Notes: Column (1) to (3): Difference- in- Differences (DD) using children aged 4 only and Column (4) to (6) Difference- in- Difference in Differences (DDD) including children aged 3 or 5. When more than one child 3 or 5 is living in the household, I consider attendance of the youngest child only. Column (1) and (4) present estimated effects for the sample of all children with no missing values for the exact date of birth. Column (2) (3) and (5) (6) are restricted samples including only children living in households where the mother is identified with no missing values on maternal labour outcomes. All regressions control for agglomerate fixed effects, survey year*wave fixed effects, child's age (by Group according to month of birth) fixed effects, a child's gender dummy and treatment time dummies. Column (4) to (6) also include second level interactions as described in the text. Additional controls include: mother's education dummies, mother's age dummies, a dummy for mother is single, total number of sisters/brothers between 0-2, 3-5, 6-13 and 14-18. Standard errors in parentheses are clustered by agglomerate_year (138 clusters). Confidence levels, ***: significant at 1% confidence level, , ** at 5% and * at 10 % confidence level. Mean of attendance for age 4 children in Cordoba during the baseline period is 0.279 and 0.312 for the restricted sample in which the mother is identified.

1.4.2 The effect of new rules on maternal labour outcomes

One way to measure the effect of new rules on maternal labour outcomes is to compute the differential change in employment of potentially affected mothers in Cordoba during the period in which the new rules were in force versus periods under the old rules and compare it to the differential change in employment of similar mothers in the provinces where no new rules had been passed. That is, using the difference-in-differences method.

By using this method, I am only able to control for time specific shocks to the labour market that are common across provinces and for time invariant province-specific levels of maternal employment. If any additional shock affects the labour market behaviour of mothers in Cordoba during the period of the implementation of these rules, then it might not be possible to properly separate this effect from the effect of the new rules. In fact, the credibility of this type of estimate is based on the assumption of parallel trend's outcomes in the absence of the new rules. This assumption could be difficult to defend given the facts observed in the analysis of the labour market described in Section 1.3.2. Effectively, these facts suggest that during the pre-treatment period a general recession affected the whole country indeed hitting heavily Cordoba compared to the rest of the country. In addition, since 1996 the recovery seems to be at a faster peace in this treated province. To deal with this potential problem, I implement two complementary approaches.

The first one, as in the previous section, is to add a third comparison group to the computation of the differences: mothers who were unaffected by the new rules because of their children's ages. This additional source of identification alleviates the potential problem because any labour market fluctuation in Cordoba affecting mothers of 4-year-olds likely has similar effects on those mothers of children of a different age. In order to make this comparison group more similar to the treatment group, and then reduce the possibility of the existence of omitted interactions, I include mothers of children aged 3 or 5. In particular, I estimate a modified version of model (1), where now i refers to the mother, the dependant variable, A_{ipt}, refers to maternal labour market outcomes discussed in Section 1.4.2. Compulsory is a dummy for being affected by the new rule of compulsory preschool, that is, if the mother resides in Cordoba during 1997-1999 and has a child aged 58 or 59 months by June 30th of the survey year, meaning that the child belongs to Age_{4Gr1}, Closure is a dummy for being affected by the new rule of preschool closure, that is, if the mother resides in Cordoba from 1996 to 1999 and has a child aged 48 to 57 months by June 30th of the survey year, meaning that the child belongs to Age_{4Gr2}. As before, I control for agglomerate fixed effects, survey year*wave fixed effects, child's age fixed effects (in this case, dummies for age 3 Gr1, 3 Gr2, 4 Gr1, 4 Gr2, 5 Gr1), a child's gender dummy and treated time dummies. In addition, I control for second level interaction dummies (Cordoba*treated time dummies; Cordoba*child's age dummies, treated time dummies*child's age dummies). Mother's controls include: mother's education dummies as discussed in Section 1.4.2, and a dummy for missing values in education variables, mother's ages dummies (7 categories: up to 22 years of age, from 23-26, 27-30, 31-34, 35-38, 39-42, 43 or above). Household's controls include: a dummy for spouse present and total number of sons/daughters living in the households between the age of 0-2, 3-5, 6-13 and 14-18.

The second approach is to find a better contrafactual for Cordoba using the/a synthetic method (Abbadie, etal., 2010). In doing so, instead of considering all other agglomerates as the control group, I take a weighted average of these agglomerates and create a "synthetic Cordoba". The aim is to choose the agglomerates that best approximate the behaviour of Cordoba's labour market outcomes in the pre-intervention period, and thus provide a better estimate of the outcomes in the absence of new rules during the treatment period. The method is particularly suitable in this case, as the new rules affected just one aggregate unit[14].

The Sample

For the main analysis, I consider a sample of mothers with the following characteristics: (a) have at least one child of preschool age (3-5 years old) with complete data on date of birth and school attendance, (b) are between 18 and 49 years old (c) have no missing data on labour market variables. This sample consists of 62,283 observations. Summary statistics for this sample are presented in Table 1.2, for Cordoba and the rest of the country, during the period of the new rules (1996 to 1999) compared to the period of the old rules (1995 and 2000 to 2001).

[14]It is important to note that I do not apply this method to the case of attendance rates for a number of reasons. First, there is no survey data available before 1995. Second, national administrative data for attendance disaggregated by Sala is only available from 1996 onwards. Third, even when I could apply the method using aggregate rates for Sala 1, 2 and 3 descriptive evidence presented in Section 1.3.1 showed that trends in attendance rates for children 3-5 were similar prior to the implementation of the new rules. Furthermore, there are no variables for the pre-intervention period that I could use as predictors of attendance.

TABLE 1.2

Summary Statistics for mothers with at least one child of preschool age

	Cordoba				Rest of the Country			
Variables	**1995/00-01**		**1996-99**		**1995/00-01**		**1996-99**	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Mother's age	32.12	6.24	32.59	6.25	32.11	6.36	32.13	6.34
At most incomplete secondary	0.59	0.49	0.58	0.49	0.62	0.49	0.63	0.48
At most incomplete terc/Univ	0.27	0.45	0.26	0.44	0.26	0.44	0.25	0.43
Complete terciary/Univ	0.13	0.34	0.16	0.36	0.13	0.33	0.12	0.33
Single mother	0.09	0.28	0.09	0.28	0.10	0.30	0.09	0.29
Total children 0-2	0.41	0.59	0.41	0.59	0.44	0.61	0.46	0.62
Total children 3-5	1.15	0.38	1.18	0.42	1.16	0.38	1.17	0.39
Total children 6-13	0.98	1.12	1.11	1.13	1.12	1.16	1.14	1.16
Total children 14-18	0.23	0.53	0.25	0.59	0.26	0.59	0.26	0.60
Prop. of other adults	0.10	0.30	0.09	0.29	0.09	0.29	0.09	0.29
Mother works	0.36	0.48	0.35	0.48	0.38	0.49	0.38	0.49
Hours of work	12.18	19.80	11.42	19.30	12.24	18.87	12.49	19.21
Hourly wages	1.26	2.50	1.12	2.09	1.35	2.74	1.41	2.86
Mother unemployed	0.07	0.25	0.06	0.24	0.06	0.25	0.05	0.22
Child is boy	0.49	0.50	0.50	0.50	0.51	0.50	0.51	0.50
Total Obs	1711		1000		20242		39330	

Source: EPH 1995-2001.

Notes: The sample consists of 62,283 mothers with at least one child aged 3-5 and non-missing mother's labour market data. Hours of work have a value of zero for mothers who did not actually work the previous week of reference. The average number of hours per week in the sample for those who actually worked is 33.6 and earning on average \$4.02 per hour. Note that 3.59 percent of those who declared having a job recorded zero hours of work.

Mothers in the sample are on average 32 years old and 10 percent of them are single. Mothers living in Cordoba are slightly more educated and have fewer children compared to mothers in the ROC. I will control for these variables in the regression analysis. The next rows present means of maternal labour outcomes. The first one – 'Mother works' - refers to the proportion of mothers who worked the previous week before being interviewed. On average, 37 percent of mothers worked. Mothers from Cordoba tend to work less than the rest of the country. The average number of hours worked last week (in all jobs) is around 12. Note that this variable takes the value of zero for all those mothers who did not work last week. Observations with more than 84 hours of work a week were considered as missing values. The average number of hours considering just the subsample of mothers who actually worked is 33.6 hours, earning on average $4.02 per hour in real terms. Finally, -'Mother unemployed'- measures the proportion of mothers who were looking for a job the week before the survey was conducted. Observations with missing values for any of these outcomes were excluded from the sample.

Regression Results

Table 1.3 reports DDD estimates for the effect of each new rule on maternal labour outcomes based on a sample of mothers with at least one child aged 3-5. The estimates in columns (5) to (8) include controls for mothers and households. I will first discuss coefficients associated with the effect of Salas closure shown in the second row.

There is no evidence of changes in labour market outcomes for the closure of Salas 4. None of the coefficients estimated are statistically different from zero.

Anticipatory behaviour could be one of the channels absorbing the effects. As was mentioned earlier, although the new rule had been suddenly announced leaving no room for parental anticipatory behaviour in 1996, reactions in the years after 1996 could have existed. If that is the case, mothers could have learnt about the new rule prior to be actually affected offsetting any observed effect on maternal labour outcomes. It could be the case that working mothers responded to the new rule by substituting public preschool provision with more informal arrangements. These informal arrangements could range from asking neighbours or friends for help to grouping with other families also affected by the policy to share the cost of private nannies. This interpretation is related to the empirical literature that shows how changes in the

TABLE 1.3

Estimates of the Effect of New Rules on Maternal Labour Outcomes

New Rules:	Dependant Variable: Mother Works (1)	Hours Worked (2)	Mother Unemployed (3)	Hourly Wages (4)	Mother Works (5)	Hours Worked (6)	Mother Unemployed (7)	Hourly Wages (8)
Compulsory Attendance	0.117	5.058***	-0.0447**	0.115	0.0429	2.284**	-0.0434*	-0.230
	(0.0767)	(1.848)	(0.0214)	(0.315)	(0.0582)	(1.156)	(0.0223)	(0.348)
Preschool Closure	0.0250	0.803	0.0154	0.163	0.0279	0.915	0.0121	0.221
	(0.0185)	(1.926)	(0.0264)	(0.105)	(0.0197)	(1.918)	(0.0253)	(0.124)
Observations	62,283	62,283	62,283	62,283	62,283	62,283	62,283	62,283
Controls	No	No	No	No	Yes	Yes	Yes	Yes

Source: Encuesta Permanente de Hogares, May 1995 to October 2001.

Notes: Column (1) presents the estimated effects for the sample of mothers with at least one child aged 3-5 with no missing values on exact date of birth and maternal labour outcomes. All regressions control for agglomerate fixed effects, survey year*wave fixed effects, child's age (by Group) fixed effects, a child's gender dummy, treatment time dummy and second level interactions. Columns (4) to (6) inclue additional controls: mother's education dummies, mother's ages dummies, a dummy for mother is single, total number of sisters/brothers between 0-2, 3-5, 6-13 and 14-18. Standard errors in parentheses are clustered by agglomerate_year. Confidence levels, ***: significant at 1percent confidence level, , ** at 5% and * at 10 % confidence level.

public financing of childcare could also have an impact on the shift in the mode of childcare (Baker, Gruber and Milligan, 2005 and Havnes and Magne Mogstad, 2009).

One way to partially test the existence of anticipation effects is to examine whether early affected mothers (i.e. those affected in 1996) reacted differently from mothers that had more time to adjust to the new policy. That is, estimating the effect of the preschool policy disaggregated by years. However, the year by year sample size for the treated group is too small to carry out this type of analysis. In particular, in 1996 only one survey round was carried out in Cordoba and in one conglomerate, restricting the size of the treated group. The same limitation applies to test heterogeneous responses across different demographic groups. Due to sample size, it is not possible to study , for instance, whether an effect of the policy could exist for single mothers, or for mothers without younger children, or for those with less years of education completed.

Changes in family composition could be another channel absorbing effects of this new rule. In addition to or instead of informal arrangements, one could speculate that families affected by the lack of public preschool could make grandparents move into their own households to take care of the children. As a result, changes in the proportion of extended families in comparison to nuclear families in which the only adults are parents should be observed. To assess this, I estimate a similar version of model (1) but use household type as dependent variable. Two different measures of household type are used: Type I and Type II. Type I takes the value of 1 if the mother or father is head of the household and there are no other adults living in the household, and takes the value of 0 in any other case - i.e. the mother or father is head of the household and there is at least one more adult living with them or households in which the head of the household is the grandfather or the grandmother or an adult but not their parents. Type II takes the value of 1 if the head of the household is the grandfather or the grandmother (of the child in question) or an adult but not her/his parents and takes the value of 0 in any other case. Again, the sample consists of all children aged 3-5, including those for whom the mother is not identified. The results shown in the second row of Table 1.9 in Appendix provide no evidence that the effect of preschool closure could have affected family composition in a significant manner. I will discuss the effect of the mandatory rule on household composition in the next subsection.

The effects of compulsory attendance on maternal labour outcomes are shown in the first row of Table 1.3. Results reveal a clear and significant effect at the intensive margin, with mothers working on average 5 hours more per week. Controlling for individual characteristics, the estimate is about half but still very significant showing an increase of 2.3 hours of work per week. On the other hand, there is a positive effect of the probability of working, although it is not statistically different from zero.

Column (3) reports the effect on the probability of being unemployed. In Column (7) I report the same coefficient in a model with the full set of controls. The negative and significant result found is somewhat surprising. Under a labour supply and demand framework, if the new rule of mandatory attendance is affecting the labour supply curve, jointly with an increase in worked hours it is expected to find, if any, an increase in the unemployment rate. Instead, what it is observed seems to be the result of a labour demand shock, with an increase in hours worked and a reduction in the proportion of mothers looking for a job.

In order to explore this issue further, I estimate the effect of the new rule on hourly wages. While under a positive labour supply shock I would anticipate a decreased (or zero) effect on wages, under a labour demand shock I would expect wages to rise. Column (4) reports the effect of the new rule on hourly wages, the coefficient found is not distinguishable from zero. In column (8), once I control for maternal and household characteristics, the coefficient is still not significantly different from zero and has a negative sign, making less likely the hypothesis that a labour demand shock is driving the results.

This latter hypothesis of a labour demand shock driving the results can be also assessed by looking at labour market outcomes of individuals who may not be affected by the new preschool attendance rule but be exposed to other shocks. Assuming that the mother is the main caregiver of the child in the household, the new mandatory rule should not have any differential effect on labour outcomes of fathers. Finding effects of a father's outcomes could be an indication of other confounding factors operating during the period of the policy change. To investigate this, I estimate the effect of compulsory attendance rule on labour outcomes of fathers. Analogous to the case of the sample of mothers, I use a sample of fathers with the following characteristics: (a) have at least one child of preschool age with complete data on date of birth and school attendance, (b) are between 18 and 59 years old (c) have no missing

TABLE 1.4

Estimates of Effect of New Rules on Fathers' Labour Outcomes

	Dependant Variable:							
New Rule:	Father Works (1)	Hours Worked (2)	Father Unemployed (3)	Hourly Wages (4)	Father Works (5)	Hours Worked (6)	Father Unemployed (7)	Hourly Wages (8)
Compulsory Attendance	-0.0212	-0.220	0.0211	0.210	-0.0158	-0.150	0.0220	0.165
	(0.0286)	(2.888)	(0.0215)	(0.321)	(0.0286)	(2.642)	(0.0228)	(0.152)
Sample Means	0.91	44.62	0.07	3.13	0.91	44.62	0.07	3.13
Observations	37,499	37,499	37,499	37,499	37,499	37,499	37,499	37,499
Controls	No	No	No	No	Yes	Yes	Yes	Yes

Source: Encuesta Permanente de Hogares, May 1995 to October 2001.

Notes: Column (1) presents the estimated effects for the sample of fathers with at least one child 3-5 (excluding children aged 48 to 57 months - Group 2) with no missing values on exact date of birth and maternal labour outcomes. All regressions control for agglomerate fixed effects, survey year*wave fixed effects, child's age fixed effects, a child's gender dummy, treatment time dummy and second level interactions. Columns (4) through (6) include additional controls: father's education dummies, father's ages dummies, a dummy for father is single, total number of sisters/brothers between 0-2, 3-5, 6-13 and 14-18. Standard errors in parentheses are clustered by agglomerate_year. Confidence levels, ***: significant at 1% confidence level, , ** at 5% and * at 10 % confidence level.

data on labour market variables. Means of dependent variable are reported in the second row of Table 1.4 showing that the proportion of fathers working is significantly higher than mothers. On average, 91% of fathers of children aged 3-5 have worked 44 hours the previous week, receiving a salary of 3.1 pesos per hour. The unemployment rate is around 7%.

Results for fathers' outcomes are reported in the first row of Table 1.4. Column (5) to (8) includes the full set of controls: father's education dummies, a dummy for missing values in education variables, father's age dummies, a dummy for father is single and the total number of sons/daughters living in the households between the age of 0-2, 3-5, 6-13 and 14-18. None of the coefficients are different from zero, providing additional evidence that the treatment effects found for the case of mothers are unlikely to be due to factors other than the new compulsory rule.

The findings thus far indicate that although closure of Salas 4 had some impact on preschool attendance, it had no effect on maternal labour outcomes. The new compulsory rule, instead, had not only an effect on preschool attendance but also one on maternal labour outcomes. That is, the first stage and reduced form show both significant effects of the new compulsory rule. As presented in the next section, the ratio between these two effects is the Instrumental Variable estimator of the effect of attendance on labour outcomes calculated using the new rule as an instrument.

1.4.3 The actual effect of attendance on maternal labour outcomes

This section investigates if attendance is the channel through which the new compulsory rule affects maternal outcomes.

Using the mandatory enrolment rule as an instrument for actual preschool attendance, the impact will be measured on those who complied with the new rule. That is, because the attendance effect is expected to vary across the population- i.e. the treatment effect is heterogeneous. This is the Local Average Treatment Effect - LATE. It is a meaningful parameter in this context as it estimates the effect of attendance on maternal labour outcomes on mothers whose children attend preschool because of the new rule, but would not otherwise have attended. Formally, LATE is defined as "the average treatment effect for individuals whose treatment status is influenced by changing an exogenous regressor that satisfies an exclusion restriction" (Imbens and Angrist, 1994 : 467)[15].

The exclusion restriction is satisfied if the effect of the treatment on the outcome is through attendance. Therefore, before examining IV estimates it is important to rule out other channels through which the new rule could affect maternal outcomes. As discussed in the previous sub-section, the new rules could change household composition which in turn affect maternal outcomes. If this is the case, then it is no longer true that the only reason why we observe changes in hours worked is because of change in attendance. Going back to Table 1.9 in Appendix in which the effect of new rules on household composition is examined, the coefficients estimated in the first row provide no evidence that the compulsory rule had any effects on these variables. As remarked earlier, the importance of this result lies in the fact that the assumption of exclusion restriction is likely to satisfy[16].

Within this heterogeneous treatment effect framework, a further assumption named "mono-

[15]Refer to Angrist and Pischke (2009), Chapter 4, Section 4 for a clear discussion on IV with Heterogeneous Potential Outcomes.

[16]See Angrist, Imbens and Rubin, (1996) for a discussion of this assumption in a context of non-compliance.

tonicity" is needed to identify LATE on these compilers (Imbens and Angrist,1994). There is no one who does the opposite of his assignment, no matter what the assignment. This is a plausible assumption in this context, as it can be expected that children aged 4 attend at least as much as they would have attended had they been under treatment.

Table 1.5 displays the LATE estimates using Two Stage Least Squares. Column (1) first row replicates the effect of the new compulsory rule on attendance (first stage) and the second to fifth rows replicate the Intention to Treat estimates from previous sections. In practice, I re-estimated these coefficients using the same sample of children aged 3-5 for whom the mother is identified but separately for each new rule[17].

In Column (2) and (3) a Two Stages Least Square (TSLS) method is used to estimate the actual effect of attendance on maternal labour outcomes. If it can be assumed that the effect of the compulsory new rule on maternal labour outcomes is produced by preschool attendance, the effect of attendance on the outcome (reported in Column 2) equals the ratio of the effect of the new compulsory rule on the outcome (reported in Column 1 second row) to the effect of the treatment on the probability of preschool attendance (reported in Column 1 first row). Under unbiased estimates of the latter two effects, this ratio will be an (approximately) unbiased estimate of the effect of a change in attendance on maternal labour outcomes.

Finally, while Column (2) uses all conglomerates as control group, Column (3) uses the "Synthetic Cordoba". That is, each observation in the sample is weighted according to the mother's agglomerate of residence. Weights are constructed using the Synthetic Control Method (Abadie, Diamond and Hainmueller, 2010). The next sub-section explains the construction of weights and displays how much each agglomerate contributes to the "Synthetic Cordoba".

Results in Column (2) show that mothers who have been induced to send their children to kindergarten are 18 percentage points more likely to work. This estimate, however, is imprecise and its effect cannot be distinguished from zero.

In the next row, instead, the effect of attending preschool on weekly hours is quite precise providing evidence that a mother works, on average, 9.17 hours more per week because her child is going to preschool. The estimate in this case is significantly different from zero at 5

[17] In this section, I only present IV results for the new Compulsory rule. Given that there was no effect of Salas closure shown on maternal outcomes, it was not relevant to display the insignificant IV estimations. Results are available upon request.

percent level.

These results are in line with those found by Berlinski, et al. (2011) who use the same source of data but apply a different research design. Their study uses a Regression Discontinuity Design exploiting the differences in the probability of attendance around the cut-off date in the year the child turns five. Despite the lack of significant results on maternal employment, they do find that mothers work on average 7.8 more hours per week because their youngest child is attending preschool. The 90 percent confident interval for this effect of attendance on hours worked ranges from 0.15 to 15.45 hours per week[18].

Finally, there is no effect of attendance on unemployment or hourly wages. Moving to Column (3) in which the control group of provinces is constructed based on synthetic methods, results are similar but suggestive of larger effects and less precise likely because of the reduction in the sample sizes. It is worth noting that hours of work is still positive and significant.

Improving the control group: Synthetic Method Synthetic method finds a set of weights that would reflect how "close" each of the control provinces is to Cordoba. The closeness is based on a set of pre-intervention (selected) variables that predicts the key outcome variable on which the contrafactual needs to be similar to Cordoba. In this way, instead of selecting provinces that a priori seem to behave in a similar way to Cordoba before the implementation of the new rules, this method constructs a "synthetic" Cordoba making explicit the relative contribution of each province (control unit) to the contrafactual of interest. I used the Synth command provided for STATA. For details about how the methods construct the "Synthetic Cordoba" as convex combinations of multiple agglomerates (control) units, see Abadie, Diamond, and Hainmueller (2011), for R package.

Using the EPH, I construct a biannual panel of agglomerates for the period 1989 to 1995 with 14 periods before the new rule comes into force. The two outcome variables of interest for which there is complete time series data available are maternal employment and maternal unemployment for mothers of children aged 3-5. The predictors used are: mother's education-proportion of mothers with low, medium and high education, proportion of mothers in different

[18] In my case, I am not able to study the effects separately for when a child is the youngest in the household or when it is not the youngest because of small sample sizes. My regressions control for the presence of youngest children in the household, among other controls.

TABLE 1.5

Estimates of the Effects of Preschool Attendance on Maternal Labour Outcomes

Dependant Variables	Compulsory rule effect estimates (1)	TSLS estimates of attendance effect (2)	(3)
Attendance	0.256***		
	(0.0817)		
Mother works	0.0461	0.180	0.278
	(0.0587)	(0.188)	(0.213)
Hours of work	2.351**	9.170**	11.54*
	(1.158)	(3.945)	(6.378)
Mother unemployed	-0.0428*	-0.167	-0.210
	(0.0240)	(0.108)	(0.167)
Hourly Wages	-0.230	-0.897	-0.232
	(0.348)	(1.482)	(1.496)

Source: Encuesta Permanente de Hogares, May 1995 to October 2001.

Notes: Column (1) first row presents the coefficient of the first stage regression. Column (1) row 2 to 5 presents the reduced form estimates of the effect of the new compulsory rule on maternal outcomes. Column (2) shows the IV estimates and Column (3) is the same estimates as in Column (3) using synthetic Cordoba as control provinces. All regressions control for agglomerate fixed effects, survey year*wave fixed effects, child's age (by Group) fixed effects and a child's gender dummy and treatment time dummy. Additional controls include: mother's education dummies, mother's age dummies, a dummy for mother is single, total number of sisters/brothers between 0-2, 3-5, 6-13 and 14-18. Standard errors in parentheses are clustered by agglomerate_year. Confidence levels, ***: significant at 1% confidence level, , ** at 5% and * at 10 % confidence level. Total number of observations in column (1) and (2) is 41,360. The smaller sample size compared to Table 1.4 arises because separately regressions for each rule were estimated in this case. The sample used by Synthetic Method is reduced to 14,957 observations and 11,642 for the case of unemployment.

TABLE 1.6
Agglomerate weights in the synthetic Cordoba

Agglomerates	**Unit_Weight Employment**	**Unit_Weight Unemployment**
GBA	0.109	0.013
La Plata	0	0.334
Rosario	0.442	0
Santa Fe	0	0.011
Parana	0.389	0
Mendoza	0.06	0.531
Neuquen	0	0
Jujuy	0	0.111
Salta	0	0
San Luis	0	0
San Juan	0	0

Source: Encuesta Permanente de Hogares, May 1989 to May 1995.

Note: GBA is Greater Buenos Aires. Rosario is an agglomerate in Santa Fe province. Santa Fe refers to Santa Fe Capital.

age groups, proportion of children under 3 years old living in the household, proportion of children aged 3-5, 6-12, 13-18, average number of other adults living in the household, share of employment in main sectors (Education, Health and Community Services, Domestic Servants and Commerce).

I use all available agglomerates with data since the late 1980s as potential controls. For the treated province, I am only able to use Cordoba Capital agglomerate, as Rio Cuarto was incorporated into the EPH from 1995. For Santa Fe province (Rosario and Santa Fe Capital agglomerates), there are missing waves in 1992. In 1993 Mendoza and San Juan do not have the variable "Sector". In all these cases, I used the averages of the neighbouring waves to impute the missing values.

Table 1.6 shows the weights of each agglomerate in the synthetic Cordoba for both outcomes: employment and unemployment. The weights reported in the table indicate that employment trends in Cordoba prior to the implementation of the new rules is best reproduced by a combination of Rosario (located in the province of Santa Fe), Parana (located in Entre Rios Province and the neighbouring city of Santa Fe), Greater Buenos Aires and Mendoza. Results for the unemployment indicate that similar to the case of employment, agglomerates from Buenos Aires (La Plata, GBA), Santa Fe (Santa Fe city) and Mendoza (Mendoza city) are the provinces with the highest contribution to the total synthetic Cordoba. In addition, Jujuy is also weighted as part of the control group provinces. All other agglomerates were assigned zero weight using this Synthetic Method.

1.5 Conclusion

This paper has been concerned with how the provision of preschool childcare affects maternal labor market outcomes. The problem of non random selection into early education has been overcome by exploiting two natural experiments that dramatically affected the probability of preschool attendance of children aged 4 in Cordoba - one of the three most important provinces in terms of population in Argentina - from 1996-99. In particular, I analysed the effect of two distinct and opposite policy changes: (a) the introduction of mandatory - free and public - kindergarten and (b) the closure of free and public kindergarten. These policies affected preschool attendance of children differently depending on their month - year of birth and province of residence. Furthermore, given the way these new rules were implemented, there is no concern on their being endogenously determined.

An historical analysis of aggregate enrolment data from the early 1980's until mid 2000 depicted an increasing trend in preschool attendance with a clear break at the time of the introduction of these policies in Cordoba compared to the rest of the country.

The effects of these policies on preschool attendance were empirically confirmed by examining the official household data - Encuesta Permanente de Hogares (EPH). Microdata from the EPH was available before and after the policies were implemented. Using the pooled repeated cross sections from all survey waves between 1995-2000 - the period for which the exact date of birth was collected - I compared preschool attendance rates of children aged 4 in Cordoba with the rest of the country during times with and without the new rules in force. This difference-in-differences estimate showed that preschool attendance has increased by about 39 percentage points due to the mandatory policy. On the other hand, because of closure of public kindergarten there was a 46.5 percent decrease of the baseline attendance rate in Cordoba relative to the rest of the country. It is worth noting that by controlling for additional individual and household characteristics, the estimates changed little while slightly gaining precision.

In an attempt to control for specific shocks affecting Cordoba during that times I also performed a triple difference-in-differences estimate. That is, I included an additional comparison group - children aged 3 or 5. The difference in attendance rates for children aged 4 between Cordoba and the rest of the country and before and after the implementation of the new rules was contrasted with the equivalent difference among children aged other than 4. The results provided evidence that estimates are statistically similar to those obtained using just children aged 4.

Having shown that these new rules differentially affected enrolment into preschool, I exploited them to identify the causal effects of provision of preschool on maternal labour outcomes. Using the new rule as an instrument for actual attendance, I measured the impact among those assigned and complying with the new rule. The parameters estimated were interpreted as Local Average Treatment Effects (LATE).

The results indicated, first, that preschool attendance had an effect on the intensive margin, mothers work on average 9 hours more per week because a child was attending preschool. This new compulsory rule would affect those who had not been attending if the rule had not been set. That is, compilers in this case were those whose probability of attendance before the law was zero and now because of this new rule attend pre-school. They might come from more disadvantaged backgrounds and are likely to live in "less traditional" households.

Second, mothers who have been induced to send their children to kindergarten are 18 percentage points more likely to work. However, this effect at the extensive margin was much less precisely estimated.

One concern with the identification strategy was that the implementation of policies happened very close to a financial crisis period. Cordoba could have been affected differently during and after the crisis compared to the rest of the provinces. In addition to a triple difference-in-difference strategy, and in an effort to create a contrafactual comparison group which better matched the treatment province in terms of pre-trend outcomes, I used the novel synthetic methods. The methods allowed me to construct a weighted combination of observations from agglomerates that behaved similar to Cordoba in terms of labour outcomes before the new rules were implemented. Using synthetic weights, the main results still hold and suggest larger effects albeit less precise likely because of the reduction in the sample sizes.

Another concern was whether the exclusion restriction is likely to hold. An analysis has been performed to examine whether the new rule could affect maternal labour supply not through attendance but through other channels, such as changes in family composition. No evidence was found that could invalidate the interpretation of estimates.

Third, when I considered the effect of closure of Sala 4, I found no evidence of changes in labour supply at the extensive or intensive margins. Mothers potentially affected by this policy might be those who would have sent their child to preschool but because of the new rule have looked for alternatives modes of care. One possible interpretation could be that these mothers responded to the new rule by substituting public preschool provision with more informal arrangements.

Finally, with the data at hand I was not able to study if stronger effects exist during the first year of the implementation or for a subgroup of mothers, such as those with younger children or single mothers. However, unlike other studies, the importance of my results lies in the fact that by studying these two policies simultaneously it was possible to learn that the effect of providing less or more preschool childcare on maternal labour outcomes is not symmetrical. That is, observing an increase in labour force participation because of an increase in attendance does not mean that a reduction in free public preschool would lead to a reduction in labour female participation.

1.6 Appendix

1.6.1 Legislative Changes in Cordoba

- Primary School Rules: Reglamento Escuelas Primarias 41009 A 38, year 1938: prohibition of children less than 5 years old (30 June) Art 23.

- Decree 8912 in 1988 (December): Changed in Art 23 Reglamento Escuelas Primarias 41009 A 38. Recognition of preschool for children of 5 years old by June 30, 4 and 3 respectively.

- Decree 1628. (October 1996): Changed Art 23 of Decree 8912, year1988. New cut-off date for preschool enrolment: only children aged 5 before August 31st of the school year. • Decree 2158. (October 1999): Back to the old system, cut-off date by 30 June.

- Decree 278. (March 1st 2000):. Reopen Salas 4.

1.6.2 Additional Tables and Figures

TABLE 1.7
Attendance rates for Sala 3 in Cordoba and ROC

	Cord			ROC*			ROC		
Year	Mean	SD	Obs	Mean	SD	Obs	Mean	SD	Obs
1995	0.02	0.13	152	0.03	0.18	3680	0.10	0.30	4276
1996	0.02	0.13	124	0.03	0.18	3863	0.09	0.29	4498
1997	0.02	0.14	192	0.02	0.15	3596	0.08	0.28	4206
1998	0.01	0.09	297	0.02	0.15	4964	0.15	0.35	5938
1999	0.01	0.11	276	0.02	0.15	4300	0.21	0.40	5206
2000	0.02	0.13	178	0.03	0.16	2465	0.22	0.41	3003
2001	0.01	0.09	163	0.03	0.17	2405	0.19	0.39	3000

Source: EPH data. Means are calculated using population weights. *ROC excluding Buenos Aires Province and City Autonomy of Buenos Aires

TABLE 1.8

Number of children enrolled in Preschool by Sector. Cordoba and Santa Fe. 1994-2003

	Cordoba			Santa Fe		
Years	Public	Private	Total	Public	Private	Total
1994	15282	4681	19963	16668	11553	28221
1995	16225	4970	21195	17237	11545	28782
1996	637	4450	5087	19170	10530	29700
1997	1283	5398	6681	19623	10525	30148
1998	465	5280	5745	20524	10843	31367
1999	216	5338	5554	22265	11336	33601
2000	22763	5595	28358	22543	11220	33763
2001	24740	7182	31922	22543	11220	33763
2002	26104	7189	33293	22543	11220	33763
2003	27625	7736	35361	24153	11172	35325

Source: 1994:Schools and Teachers National Census 1994. Córdoba 1995: Mensaje Apertura Sesiones Legislativas. Dirección General Red Federal de Información Educativa. 1996-2004. Santa Fe 1994-2003 Dirección General de Información y Evaluación Educativa del Ministerio de Educación. Data for Santa Fe 1995 is aggregated by child's year of age. Dissagregation presented is based on assuming same trend for subsequent years.

TABLE 1.9

Estimates of the Effects of New Rules on Household Composition

	Dependant Variable:	
New Rules:	Type 1 HHs (1)	Type 2 HHs (2)
Compulsory Attendance	-0.0104	-0.00148
	(0.0521)	(0.0431)
Preschool Closure	0.0438	-0.0445
	(0.0558)	(0.0409)
Mean (baseline)	0.790	0.098
Observations	95,782	95,782
R-squared	0.028	0.018

Source: Encuesta Permanente de Hogares, May 1995 to October 2001.

Notes: The sample consists of all households with at least one child aged 3-5 with no missing values on exact date of birth. Type I HHs equals 1 if the child aged 3-5 is son/daughter and there is no other adult living in the household except their parents or mother or father, and equals 0 in any other case. Type II HHs equals 1 if the child 3-5 is grandson/granddaughter or has any other relationship different from son/daughter with the head of the household, and equals 0 otherwise. Preschool Closure equals 1 if the household is affected by the new rule of Closure of Salas4. Compulsory Attendance equals 1 if the household is affected by the new rule on Compulsory Attendance. Regressions control for agglomerate fixed effects, survey year*wave fixed effects, child's age (by Group) fixed effects, treatment time dummy and second level interactions. Standard errors in parentheses are clustered by agglomerate_year (138 clusters). Confidence levels, ***: significant at 1% confidence level, , ** at 5% and * at 10 % confidence level. Mean (baseline) refers to mean of the dependant variable in the year of 1995.

References

1. Abadie, A., Diamond, A. and Hainnueller, J., (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. Journal of the American Statistical Association. June, Vol. 105, No. 490, Applications and Case Studies.

2. Abadie, A., Diamond, A. and Hainnueller, H., (2011). Synth: An R Package for Synthetic Control Methods in Comparative Case Studies. Journal of Statistical Software. Vol. 42, Issue 13, June.

3. Angrist, J. D., Imbens, G. W., and Rubin, D. B.,. (1996). "Identification of Causal Effects Using Instrumental Variables." Journal of the American statistical Association 91(434): 444–55.

4. Angrist, J. D., and Pischke, J. S., (2009). Mostly Harmless Econometrics: An Empiricist's Companion. Princeton, NJ: Princeton University Press..

5. Abratte, J. P., (2007). Hegemonia, Reformas y Discurso Educativo en la Provincia de Cordoba (1984-1999). Doctoral Thesis. Facultad Latinomaericana de Ciencias Sociales - FLACSO, Sede Académica Argentina.

6. Baker, M., Gruber, J. and Milligan, K. (2008). "Universal Childcare, Maternal Labour Supply, and Family Well-Being." Journal of Political Economy 116(4):709–45.

7. Berlinski, S, Galiani, S and McEwan, P. J., (2011). Preschool and Maternal Labour Market Outcomes: Evidence from a Regression Discontinuity Design, Economic Development and Cultural Change, University of Chicago Press, Vol 59(2), pages 313 to 344.

8. Bertrand, M., Duflo, E. and Mullainathan. S. (2004). "How Much Should We Trust Differences-in-Differences Estimates?" Quarterly Journal of Economics 119(1): 249–75.

9. Cascio, E. (2009). "Public Preschool and Maternal Labour Supply: Evidence from the Introduction of Kindergartens in American Public Schools." Journal of Human Resources 44(1):140–70.

10. Cameron, A. C., and Trivedi, P. K., (2005). Microeconometrics: Methods and Applications. New York: Cambridge University Press. Havnes, T. and Magne Mogstad, (2009). Money for Nothing? Universal Child Care and Maternal Employment. IZA Discussion Paper No. 4504.

11. Havnes, T. and Magne Mogstad, (2011). No child left behind: Subsidized child care and children's long-run outcomes. American Economic Journal: Economic Policy, 3(2), pp. 97-129.

12. Holland, P. W.,Statistics and Causal Inference. Journal of the American Statistical Association, Vol. 81, No. 396. (Dec. 1986)pp. 945-960.

13. Fitzpatrick, M. (2010). "Preschoolers Enrolled and Mothers at Work? The Effects of Universal Pre-Kindergarten." Journal of Labour Economics 28(1):51–85.

14. Gelbach, J. B. (2002). "Public Schooling for Young Children and Maternal Labour Supply." American Economic Review 92(1):307–22.

15. Ravallion, Martin, Galasso, E., Lazo, T. and Philipp, E. Do Workfare Participants Recover Quickly from Retrenchment? The World Bank, September 2001, Policy Research Working Paper, 2672.

16. Meyer, B. D. (1995). Natural and Quasi-Experiments in Economics. Journal of Business & Economic Statistics, Vol. 13, No. 2, , JBES Symposium on Program and Policy Evaluation. (Apr. 1995) pp. 151-161.

17. Schlosser, A. (2005). "Public Preschool and the Labour Supply of Arab Mothers: Evidence from a Natural Experiment." Hebrew University of Jerusalem, Jerusalem. Unpublished. Wooldridge, J. M. (2002). Econometric Analysis of Cross Section and Panel Data. Cambridge, MA: MIT Press.

2 The Effects of Individual Characteristics on the Distribution of College Performances

2.1 Introduction

Considerable space has been awarded in the social and human sciences to the question of how individual characteristics impact on educational performance. A quantification of how student-specific factors affect educational success is crucial to explain disparities in educational achievements, and to design and evaluate specific actions aimed at promoting upward social mobility. This requires accurate empirical models that link educational performance to its observable determinants.

As has been well documented, available models are still far from this goal, mostly due to the inherent complexity of the problem which is usually reflected in their very poor goodness-of-fit performance. For example, Betts and Morell (1999) in a related study for the University of California at San Diego, obtain R^2 coefficients around 0.15, using a rich data base of 5,623 students. This means that even after conditioning on many observable aspects that determine success, individuals still differ substantially due to the importance of unobserved factors. Consequently, the correct way to assess the effect of an observable variable on performance is to think about how changes in this specific factor affect the *conditional distribution* of performances.

As a simple example, consider the effect of the father's education. The distribution of performances conditional on observed factors, including father's education, still presents substantial variability due to the non-trivial role played by unobservables, hence, even within a group of individuals with the same observed characteristics, we will find students with bad, ordinary, or good educational performances. It is natural to expect that the whole conditional distribution of performances gets shifted to the right when, all else being equal, we consider students with better educated fathers. In the extreme case where additional father's education shifts the whole conditional distribution to the right without altering its shape, the effect of increasing father's education on the *mean* performance captures everything there is to know. In such context, and under some simplifying assumptions, a standard regression model gives the desired answer: the coefficient of fathers's education in a linear regression captures the effect on expected performance and, under these circumstances, on performance in general. This situation naturally arises when father's education is independent of non-observables in the determination of performances, hence, movements in father's education imply pure location shifts of the conditional distribution of performances. But given the non-trivial role played by unobservables in these models, we cannot discard the possibility that movements in father's education interact with factors not included in the model in a non-obvious way. It might be the case that the father's education plays a more important role in children less inclined towards study, and has only a mild effect on those more motivated. In this case, the 'mean effect' of father's education is positive but does not represent anybody in the population: it overestimates the effect on individuals with a high propensity towards study, and underestimates the situation of the less motivated students.

A second example, derived from the empirical results of this paper, is the following. Consider the effect of age on college performance. Older students may be a mix of more focused and mature individuals together with badly motivated students who have advanced slowly in the educational process. Unless we can control for abstract, difficult to measure factors like 'focus', 'maturity' and 'motivation', all else being equal, the cluster of older students might perform on average like the group of younger students, even though the former is certainly more disperse in their performance. In this case the conclusion that 'on average' age does not have an effect on performance might lead careless observers to the erroneous conclusion that age does not have any effect, ignoring its impact on the dispersion.

The main goal of this paper is to measure the effect of observable individual characteristics on the whole conditional distribution of performances using recent *quantile regression* methods. We are not trying to isolate the causal effects of the observable variables included in the model, but instead our focus is on the differences between the incremental effects of the variables at the different quantiles of the conditional performance distribution. Some of the variables under study, such as student works, can be endogenous. We do not address the endogeneity here. Furthermore, to the extent that these variables reflect remaining unobserved characteristics the estimated effects are not consistently estimate.

There are three main reasons why the quantile regression approach is relevant. First, it complements standard 'educational production function' studies by exploring effects beyond those on the conditional mean. This is important since educational policies are usually expected to have an impact on those students who face relatively more difficulties, so extrapolating the effect of the average individual may induce considerable biases in the assessment of such policies.

Second, mean effects are seldom informative about the *distributive* impacts of policies. The presence of heterogeneous effects suggests that changes in specific characteristics may have the effect of improving everyone's performance but also of altering the shape of the distribution of these performances. Quantile methods provide an informative picture of these distributive effects. This is a particularly relevant issue since education is explicitly seen by many social actors as an active equalizing policy[19]. For example, and as a preview of some empirical results, having attended a private secondary school (as opposed to state) has a positive effect on performance for individuals around the center of the distribution of their unobservable factors, and no effect for those extremely good or bad. Then, having attended a private secondary school makes the distribution of performances more *asymmetric*, a subtle but relevant effect improperly summarized by the 'average' effect.

Third, distributive results become crucial when the *screening* perspective of schooling is emphasized. In this context, what matters is the capacity of the educational system to provide information about the relative abilities of individuals, and hence, as stressed by Hanushek in his classic survey '... more attention should be directed toward the distribution of observed educational outcomes (instead of simply the means)...' Hanushek, 1986, pp. 1153.

Analysis of the relationship between educational outcomes and observed factors has been investigated more intensively at the elementary and secondary levels, leaving ample room for contributions aimed at the higher education level. The empirical application of our paper ex-

[19]The 1994 reform to the Argentine Constitution explicitly states as an obligation of the Congress to 'pass laws that assure equality of opportunities and possibilities without any discrimination; and guarantee principles of free-tuition and equity in public education as well as the national university's autonomy (inc.19)'. Moreover, among the objectives of the Higher Education Law (24521/95 - Article 4) are 'To deepen the process of democratization of Higher Education, to contribute to an equal distribution of knowledge and to ensure the equality of opportunities' (inc. e). The same article sets forth as another objective those already provided by the Federal Educational Law (No 24195/93, 5th Article) 'the achieving of effective equality of opportunities and possibilities for all inhabitants and the rejection of any kind of discrimination' (inc. f) and 'Equality through the fair distribution of educational services in order to achieve the best quality and equivalent results deriving from a heterogeneous population' (inc.g). In the same vein, the Financial Education Law (26075/05) establishes an increase in the allocation of fiscal resources for the education system up to 6% of the total GDP by 2010. Its objective is clearly stated in the first article '..to guarantee equality of opportunities of learning' (Art.1st).

ploits a comprehensive census data set that covers all students attending public universities in Argentina in 1994. Public higher education in Argentina has been operated as a free and unrestricted access system - in general without entrance examinations - during most of the last twenty years, providing a unique source of sampling variability. Students entering university come from diverse socio-economic backgrounds. Additionally, in some popular degree programs like Accountancy or Law, university rules are very flexible and leave the rate at which student´s progress up to themselves. This results in very different performances among students of the same cohort, adding greater value to our data set. To the best of our knowledge, this is the first paper that takes advantage of these particular features.

There are some relevant antecedents to our study. Eide and Showalter (1998) and Levin (2001) used quantile methods to study how school characteristics affect performance. Betts and Morell (1999) conducted a comprehensive analysis of the determinants of college performances using a large sample of University of California at San Diego students. A relevant conclusion of their paper is that '... variations in family background and in the socio-economic environment of the school play far more crucial roles in determining student outcomes in university than do variations in school resources', which is in line with empirical results that go back to the 1966 Coleman Report. Consequently, and in light of these results, we emphasize the use of quantile methods to study how *individual* characteristics impact on performance. As previously mentioned, the literature on higher education and performance is relatively scarce compared with that related to elementary and secondary education. A representative study of the literature that is available is Naylor and Smith (2004), who study the determinants of college performance for the United Kingdom and Di Gresia et. al. (2007), Porto (ed). (2007) for the case of Argentina.

The rest of the paper is organized as follows. The next section discusses the econometric strategy used to recover the effects of observed variables on conditional distributions, while linking this study to previous literature on the subject. Section 3 presents the data set utilized in the empirical part and details the particular aspects of the Argentinean system that are relevant for the purposes of this paper. Section 4 presents the econometric results, and Section 5 concludes.

2.2 Exploring distributive effects through quantile regressions

As advanced in the Introduction, non-trivial distributive effects of observed factors arise when they interact with non-observables. This section presents a simple structure for these interactions and proposes the use of quantile regressions to model them.

2.2.1 Interactions between observed and unobserved factors

The educational *production function* approach, originated in the famous 'Coleman Report' and reviewed extensively by Hanushek (1986), models educational performance as the outcome of transforming 'inputs' into educational 'outputs' in a production function fashion[20]. A stringent empirical limitation of these models is that the vector of inputs include a myriad of unobserved individual specific factors which may play a non-trivial role. To the point, in his landmark paper Hanushek (1979) states that '... the most consistent and obvious divergence of the empirical models from the conceptual models is the lack of measurements for innate abilities'. In educational production functions, these abilities play a role similar to that played by 'entrepreneurial factors' in standard micro-theory production functions, in the sense that they represent unobserved factors that imply different profits for different firms (individuals, in the case of education) which may alter the way observed factors affect production.

Consider a simple, individual specific, production function

$$y_i = g_i(x)$$

that represents the maximum educational outcome y that an individual i may produce with inputs x. In general this function is not homogeneous of degree one since each person has her/his own set of fixed 'innate abilities'. As clearly stressed in the standard microeconomic theory (i.e., Mas Collel, et al., 1995, pp. 134-35), production functions reflect technologies, not limits on resources, hence the individual specific production function is better represented

[20] The importance of understanding education as a production process had already been put forward by Olivera in 1964, stating that 'education, in some sense, is a branch of production. As well as in any other industry, (education) utilizes material and human resources and labor and capital in order to obtain some outputs' (pp.103 see also, Olivera (1967) and Araoz (1968))

by

$$y_i = g(x, u_i)$$

where u represents unobserved factors that, once fixed at a particular level, describe how x is transformed into y for a particular person.

As well known in the literature, u is far from playing a minor role in explaining educational disparities. Even when the dimension of x is large so as to include a multitude of individual and institution specific factors, u still contains abundant unobserved information about the psychological and motivational characteristics that make individuals differ in their performance[21]. Hence it is risky to proceed by making strong assumptions about the role played by u in the production process.

In particular, we are concerned with the possible interactions between unobserved and observed factors. This is a question related to the specific form of $g(x,u)$. Using a traslog specification, Figlio (1999) explores interactions among *observed* factors by testing the statistical significance of interactive terms. But if the interest is in exploring interactions between observed and *unobserved* factors we cannot rely on such a strategy.

2.3 The quantile regression approach

Consider the following general, possibly non-separable, production function:

$$y = g(x, u)$$

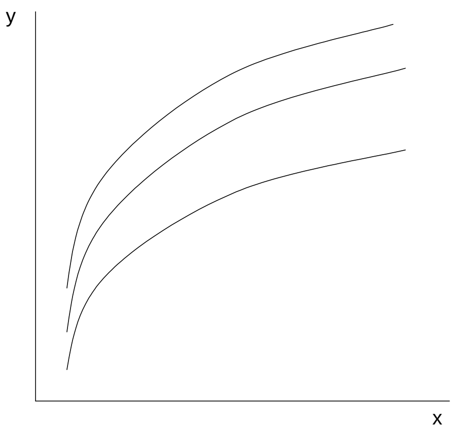
and, to simplify notation, consider the case where x is a single observed explanatory factor. Our interest is in exploring whether $\partial y/\partial x$ varies with different levels of u, and 'separability' or 'no-interaction' means that this derivative is constant across the different levels of u. Since u is not observed by the analyst it is awkward to speak about its absolute levels. Instead, it seems more convenient to consider a standardized relative notion, like its quantiles, that is, levels of u that are deemed as 'high' based on the (relative) notion that a large proportion of its possible values lie below them[22].

[21] This is typically reflected in the low goodness-of-fit performance of these models, even when large scale data bases or flexible forms are used. (Betts and Morell (1999), among others).

[22] Roemer (1998, pp.10) adopts a similar relative characterization when he uses *centiles* to measure effort.

FIGURE 2.1

Production Functions for Different Conditional Quantiles



In this context the starting point is how much can be produced of y when u is set at its τ-th conditional quantile given x, that is,

$$g(x, Q_{u|x}(\tau)),$$

where the notation $Q_{z|x}(\tau)$ stands for the τ-th quantile of the distribution of a random variable z conditional on x[23]. If, as it is standard in any production function, we assume that $g(x,u)$ is monotonic in u when x is fixed, and since quantiles are equivariant under monotonic transformations (i.e., $Q_{h(z)}(\tau) = h(Q_z(\tau))$ for any monotone function $h(.)$), then

$$g(x, Q_{u|x}(\tau)) = Q_{g|\,x}(g(x,u))(\tau) = Q_{y|\,x}(\tau). \tag{1}$$

Hence how much can be produced when u is set at any relative level measured by its conditional quantiles coincides with the τ-th conditional quantile of performances. In the simple case where x is a single production factor, in the space (y,x) this corresponds to a *family* of production functions indexed by the conditional quantiles of u. Figure 2.1 illustrates this point where each curve corresponds to $g(x, Q_{u|x}(\tau))$ for increasing levels of τ.

Consequently, our measures of interest are the partial derivatives of these production functions for different relative levels of u, $\partial Q_{y|\,x}(\tau)/\partial x$. Quantile regression models are specifically designed to estimate these derivatives.

[23]This argument follows Chesher (2003) who studies the identification of general non-separable functions.

Using the chain rule in (1)

$$\frac{\partial Q_{y|\,x}(\tau)}{\partial x} = \frac{\partial g(x,u)}{\partial x} + \frac{\partial g(x,u)}{\partial u}\ \frac{\partial Q_{u|x}}{\partial x}\ .$$

In this context ‘separability’ means that $\partial g(x,u)/\partial x$ does not depend on the levels of u and that $\partial Q_{u|x}/\partial x$ is zero, that is, u is independent of x.

As an illustration consider the linear case

$$y = \delta_0 + \beta_1 x + u,$$

so

$$Q_{y|\,x}(\tau) = \delta_0 + \beta_1 x + Q_{u|\,x}(\tau).$$

If u is independent of x

$$\frac{\partial Q_{y|\,x}(\tau)}{\partial x} = \frac{\partial y}{\partial x} = \beta_1,$$

a constant, since there is no explicit interaction between x and u and since u and x are independent.

The simplest specification that allows for these type of interactive effects is provided by the standard linear quantile regression model

$$g(x, Q_{u|x}(\tau)) = Q_{y|x}(\tau) = \alpha(\tau) + \beta(\tau)x$$

where $\beta(\tau)$ is any function. This implies that for any fixed τ, $g(x, Q_{u|x}(\tau))$ is a linear function with slope $\beta(\tau)$. Interactions arise due to the fact that for any given x the slope of these lines is allowed to vary across the conditional quantiles of u. The null hypothesis of ‘no-interaction’ corresponds to the case where all slopes are equal, $H_0 : \beta(\tau) = \beta_0$.

Based on a sample $(x_i, y_i), i = 1, \ldots, n$ of independent, though not necessarily identically distributed observations, coefficients of this model are estimated for several quantiles using the standard Koenker and Bassett (1978) estimator that solves:

$$\hat{\beta}(\tau) = \text{argmin} \sum_{i=1}^{n} \rho_\tau(y_i - x_i' b(\tau)),$$

$\rho_\tau(z) \equiv z\,(\tau - I(z<0)),\ \tau \in (0,1)$. Basic inference, such as individual significance tests and confidence intervals, is handled as follows. We estimate the vector of unknown coefficients for a grid of M equally spaced quantiles $\tau_m,\ m = 1,\ldots,M$. Let $\hat{\beta}(\tau_m)$ be each of these vectors of coefficients, and let $\beta(\tau_m)$ be their population counterparts. Collect all coefficients for all chosen quantiles as $\hat{\beta} \equiv (\hat{\beta}(\tau_1)' \cdots \hat{\beta}(\tau_M)')'$ and $\beta \equiv (\beta(\tau_1)' \cdots \beta(\tau_M)')'$. Under the assumption of a random and independent sample (not necessarily identically distributed) and under standard regularity conditions,

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{d} N\Big(0, V_n\Big)$$

where V_n is an $MK \times MK$ block diagonal matrix with blocks:

$$V_n(\tau_m, \tau_j) = [\tau_m \wedge \tau_j - \tau_m \tau_j]\, H_n(\tau_m)^{-1} J_n H_n(\tau_j)^{-1}$$

with $J_n = (1/n)\sum_{i=1}^{n} x_i x_i'$, $H_n(\tau) = \lim_{n\to\infty} \sum_{i=1}^{n} x_i x_i'\, f_i(F_i^{-1}(\tau))$, and f_i stands for the conditional density of y_i. $H_n(\tau)$ is estimated using the Koenker-Hendricks procedure. We refer to Koenker (2005) for further details.

Linear hypothesis of the type $H_0 : R\beta_0 - r = 0$, where R is a $q \times MK$ matrix and r a q vector, can be evaluated through the statistic:

$$T_n = n(R\hat{\beta} - r)'[RV_n^{-1}R']^{-1}(R\hat{\beta} - r)$$

which is distributed as $\chi^2(q)$ asymptotically under H_0, where q is the rank of R. This allows several configurations like individual or joint significance of variables, or the 'homogeneity' assumption that coefficients are equal across quantiles.

Consider the homogeneity assumption $H_0 : \beta(\tau) = \beta_0, \tau \in (0,1)$ where all slopes are equal across *all* quantiles. The previous approach handles this hypothesis through evaluating it at a discrete grid of selected quantiles. Koenker and Xiao (2002) propose appropriate tests for this hypothesis along a continuous range for τ. The homogeneity null is usually referred to as the 'pure location shift' hypothesis, since under it variables have the effect of shifting the whole conditional distribution without altering its shape. A less drastic hypothesis is the pure 'location-scale' hypothesis which implies a particular form of heterogeneity where variables shift the conditional distribution while altering its scale in a simple 'heteroscedastic' fashion. We implement both tests, and refer to Koenker and Xiao (2002) for technical details.

2.4 The Data and main features of the Higher Education System in Argentina

We base our study on the CEUN (Censo de Estudiantes de Universidades Nacionales), a national census data set that covers all college students enrolled at the 31 national universities in Argentina in October 1994, amounting to approximately 615,000 students. This data base includes detailed information on several individual and household socio-economic characteristics, as well as college performance for each student in the sample.

Because of its distinguishable institutional features, the sampling variation present in our data is ample, offering a unique empirical opportunity to study the determinants of college success. Specifically, the system of public universities in Argentina has a long tradition of promoting equality of opportunities by providing free and unrestricted access to higher education. Free tuition remained even after the Higher Education Law was passed in 1995, which provided universities with full autonomy over their administration, internal resource allocation, staff management, and student access. According to this law, it is up to the universities to decide whether or not they want to charge fees[24]. Nevertheless, the great majority of universities do not charge tuition and, in general, there are no limiting entrance examinations.

Our data is also less vulnerable to the negative effects of the selection mechanisms present in most universities, especially those in the American or British systems where strong competitive schemes determine access. For instance, in the US or the UK, students must achieve a minimum score at secondary education level in order to be considered for entry a specific university. This restriction influences the allocation of students to particular universities, biasing correlations between educational performance and students in such samples.

Other countries in Latin America or Europe (for example, Germany) also promote free access by keeping tuition levels at relatively low or zero cost, but the case of Argentina is important since free tuition is, in general, simultaneously combined with free entrance, independent of previous performance in secondary school or pre-examinations.

Compared to other countries, higher education coverage in Argentina ranks among the

[24]Higher Education Law, Chapter IV, Section 1, Article 50.

highest. According to the official Permanent Household Survey of May 2003, 65% of young people aged 18-29 years old who completed secondary school started university, and 20% of them are in the lowest quintile of the equivalized household income distribution[25] compared with 32% in the highest quintile, showing that beneficiaries of public university education come from families located in different parts of the income distribution.[26]

Our dataset provides evidence that students attending higher education come from a diverse socio-economic background and constitute a heterogeneous group. Moreover, additional statistics for a subgroup of our data show that poor people do not only start university but they also obtain a degree albeit with lower chances than those who come from richer families. For instance, following a cohort of Accountancy students at National University of Rosario from 1991 to 2001, Giovagnoli (2005) noticed that 20.4% of those who graduated had fathers with primary education or less.

An additional characteristic that makes Argentina's public higher education system an attractive case is that students in most programs have highly flexible schedules and regulations that allow them to proceed at their own pace. Hence a cohort of students could advance very dissimilarly along its academic path without being penalized, leaving ample room for individual characteristics to play a role as determinants of performance. As we will show later, data confirms this pattern. Specifically, for reasons explained below, focusing on a cohort of students who started university in 1991 and measuring their performance by 1994, we observe high variation.

2.4.1 The Measure of Performance

The choice of a particular measure of performance is a delicate issue, subject to much debate. Authors draw on measures such as GPA's (Betts and Morrell, 1999), the Graduate Record Examination (GRE) (McGuckin and Winkler (1979)) or the estimation of potential incomes (Card and Krueger (1996)), but these are not available in the CEUN data set. Nevertheless,

[25]Equivalized income takes into account the fact that food needs are different across age groups - leading to adjustments for adult equivalent scales - and that there are household economies of scale.

[26]Gonzalez Rozada and Menendez (2004) suggest that the opposite conclusion holds. They conclude that poor students tend to be excluded from higher education and hence do not obtain the benefits of free access. Their results, however, arise from comparing individuals attending higher education versus those in the relevant age group who do not attend college, regardless of their secondary school status.

there are no theoretical or empirical reasons to consider that one indicator dominates the others; in fact, there is a vast literature revising the weaknesses and the strengths, supporting the use of different measures - for a rich and extensive discussion on the issue see Hanushek (1979, 1986). In this paper we measure performance as the number of courses passed from the beginning of the program (as we will be working with the cohort of students enrolled in 1991 observed in 1994, we will measure number of courses passed after 4 years of study). In the context of understanding education as a production process as discussed in Section 2 this is a measure of average productivity. Thus, a student who passes more courses per year demonstrates greater productivity - i.e. has a better performance - than another who contemporaneously started university and passed less courses. The former student will be able to incorporate human capital in a shorter period of time leading to earning income at an earlier point in the life-cycle.

With respect to the choice of explanatory variables, the underlying theory is not explicit about any particular choice, hence data availability has played an important role in this decision. Following previous research (Hanushek (1979, 1986), Naylor and Smith (2004) and Betts and Morell (1999)), we will consider the inclusion of particular variables which can be grouped into four main types: (i) the student's demographic variables (gender, age); (ii) the student's family background (parents' education); (iii) the student's chosen factors such as the decision to work or not, city of residence/or commuting and marital status and (iv) type of school that the student attended prior to enrolling in university (public or private secondary school, type of orientation - commercial or other[27]).

2.4.2 The Sample

Our empirical analysis is focused on the two most popular programs, Law and Accountancy, at the four largest universities: University of Buenos Aires (UBA), National University of Cordoba (UNC), National University of Rosario (UNR) and National University of La Plata (UNLP). The choice of this particular sub-sample is based on the usual trade-off between increased information and heterogeneity: more programs and universities provide more sample

[27]Commercial/Administrative orientation includes basic theoretical and practical concepts about Business Administration, Accountancy and Economics. Other orientations are: (a) Humanistic, which includes concepts about different areas, specially designed to continue tertiary/university levels and (b) Technical orientation which is focused on the production process in different sectors of the economy, and is much less popular than the other two.

points at the potential cost of introducing heterogeneities between schools and programs that may obscure the goals of our analysis. Students at these four universities constitute more than 50% of total enrolment at national universities in the country and, of these, approximately 30% study Accountancy or Law (among 900 other career options).

Interestingly, these programs are quite homogeneous amongst the national universities of the country because an important part of their syllabi is related to national laws and codes, and their professional practice is subject to strict regulations[28]. This allows us to pool observations from different universities and increase precision without introducing heterogeneities at university level[29].

From the total number of individuals studying Accountancy and Law in the 1994 Census of University Population, we focused on the approximately 8.000 students who enrolled during 1991. As these programs have a nominal length of at least five or six years, the very good students of this cohort were in their fourth year at the time of the census. More recent cohorts (those who entered after 1991) had passed less courses at the time of the census, hence their measure of performance is a less precise indicator. In extreme cases, some members of the 1994 cohort had passed a few number of courses, thus the average number of courses passed may be a very poor predictor of overall performance. On the other hand, older cohorts may not be correctly represented as their best students may have finished college in less than the expected five years of study and are naturally not present at the moment the census was conducted.

2.4.3 Descriptive Statistics

Table 2.1 shows descriptions and summary statistics for the variables included in our dataset. The performance indicators reveal that after four years students in Accountancy programs had passed, on average, 12.13 courses, hence the average productivity is around three courses passed per year. Note that in the case of Law, the nominal duration of the program is one

[28] Unlike the US system, the professional practice of lawyers and accountants in Argentina requires an undergraduate degree in the subject. For example, to become a professional accountant a student must obtain an undergraduate degree in Accountancy and then obtain a professional license in the province where she/he wishes to practice her/his profession. The license is awarded automatically to every graduate, without exams, since it is understood that professional evaluation has already taken place at university. Accountancy norms are quite homogeneous across different provinces. The case of lawyers is similar.

[29] When we analyzed the structure and syllabi for each program among universities for 1991 cohorts, UNLP seems to be less flexible (especially in Accountancy) than the others. UNLP also has a greater number of missing data in the data set.

TABLE 2.1
Variable Description and Summary Statistics. Accountancy and Law sample - Cohort 1991

Name	Description	Accounting			Law		
		Mean	Percentile		Mean	Percentile	
			5%	95%		5%	95%
Performance Indicators							
performance	Number of courses passed	12.13	3	21	10.37	2	18
Explanatory Variables							
male	1 if male; 0 otherwise	0.50	0	1	0.41	0	1
age	Age in years	21.98	21	25	23.17	21	33
private	1 if private secondary school	0.49	0	1	0.44	0	1
commercial	1 if commercial secondary school	0.65	0	1	0.33	0	1
workrelated	1 if job related to program	0.41	0	1	0.22	0	1
worknotrelated	1 if job not related to program	0.26	0	1	0.40	0	1
workno	1 if student does not work	0.33	0	1	0.37	0	1
single	1 if single	0.95	1	1	0.89	0	1
cityuniv	1 if live in school area	0.75	0	1	0.74	0	1
citychange	1 if relocated	0.20	0	1	0.27	0	1
educparents	Parental education†	12.54	7	18	13.11	7	18
educfather	Paternal education	11.32	3.5	18	12.04	3.5	18
educmother	Maternal education	11.05	3.5	18	11.68	3.5	18
Dummies for Universities							
uba	1 if UBA	0.49	0	1	0.56	0	1
unc	1 if UNC	0.19	0	1	0.22	0	1
unlp	1 if UNLP	0.14	0	1	0.09	0	1
unr	1 if UNR	0.17	0	1	0.12	0	1
Number of observations		3816			4812		

Source: CEUN 1994

† Maximum between father and mother education. Mother or father education when only one of them is present.

year more than Accountancy, although the number of courses is similar, thus, by 1994 Law students would have passed fewer exams than those in Accountancy.

While the same proportion of males and females attend Accountancy, females are slighly in the majority in the Law programs (59%). The latter sample has slightly older and non-single students, and 56% of them come from a public secondary school. In Law a significantly smaller proportion of students than in Accountancy previously attended a secondary school with commercial orientation compared with those who attended a different secondary school orientation (33% versus 65%).

Labor market variables suggest a very dissimilar composition in each program related to the kind of job students have. Although around 35% of both accounting and law students said they did not have a job by 1994, working groups are different between the programs. While 41% of accounting students have a job related to their careers, only 22% of Law students have a job linked to their profession.

Regarding the characteristics of students' fathers, in Accountancy, they have, on average, 11 years of formal education, which corresponds to incomplete secondary school. Fathers of students in Law are slightly more educated. Finally, the proportion of students in each university highlights the relevance of UBA in the total sample - 56% (49%) of Law (Accountancy) students are from UBA, with UNC being the second largest in terms of the students in the sample.

2.5 Estimation Results

We have estimated a basic linear quantile regression specification using the log of performances as explained variable, for Accountancy and Law students respectively, pooling the information of the four universities considered, and including dummy variables by universities.

Tables 2.2 and 2.3 present estimation results for Accountancy and Law separately.

The first five columns of each table present point estimates of the coefficients of the linear quantile regression model of (log) performances for quantiles 0.1, 0.25, 0.5, 0.75 and 0.9. The sixth column presents standard OLS estimates which measure mean effects. The last two columns present the Koenker-Xiao statistics for the null hypothesis that the effect of each variable is a location shift and a location-scale shift, respectively. In the bottom of these two columns we present the test statistics of the global hypothesis of location shift and location-scale shift.

These tests strongly reject the null of homogeneity or pure location effects, stressing our initial point that the effect of observed factors is heterogeneous across the quantiles of unobserved factors, suggesting the presence of non-trivial interactions, in the sense discussed in Section 2.2.

Figures 2.2 and 2.3 present these results graphically. Each small picture presents the effect of each explanatory variable on the τ-th quantile of the conditional distribution for a finer grid of quantiles ($\tau = 0.1, 0.11, \ldots, 0.89, 0.9$). The solid line shows the effect at each quantile and

TABLE 2.2

Quantile Regression Results. Accountancy.

	Quantile					OLS	(a) THn. Null Hypothesis :	
Variable	0,10	0,25	0,50	0,75	0,90		Location	Location/Scale
(Intercept)	1,687	2,057	2,418	2,649	2,578	2,318	1,406	1,546
	3,888	8,945	12,288	18,330	16,995	17,141		
single	0,250	0,136	0,071	0,018	0,044	0,079	2,026	1,008
	3,552	1,609	0,910	0,336	3,433	1,526		
male	-0,177	-0,058	-0,059	-0,021	-0,012	-0,060	1,150	1,887
	-3,101	-1,922	-2,916	-1,610	-1,354	-3,001		
age	-0,023	-0,014	-0,007	-0,002	0,009	-0,008	1,776	2,442
	-1,293	-1,750	-0,978	-0,306	1,393	-1,891		
educparents	0,038	0,031	0,020	0,013	0,008	0,022	2,321	2,670
	5,559	8,092	7,579	7,885	6,948	8,824		
workrelated	-0,184	-0,114	-0,050	0,001	0,006	-0,084	1,335	2,943
	-2,608	-3,144	-2,294	0,105	0,584	-3,488		
worknotrelated	-0,356	-0,390	-0,243	-0,115	-0,077	-0,254	3,439	3,518
	-5,298	-10,108	-7,652	-5,801	-5,806	-9,562		
commercial	0,048	0,042	0,006	-0,005	-0,004	0,014	1,488	1,022
	0,860	1,262	0,293	-0,361	-0,404	0,641		
private	0,023	0,099	0,080	0,034	0,015	0,062	2,028	1,750
	0,417	3,231	3,850	2,629	1,656	3,065		
cityuniv	-0,045	-0,013	0,044	0,010	0,031	0,000	1,289	1,346
	-0,664	-0,349	1,769	0,637	2,979	0,015		
citychange	0,108	0,107	0,079	0,052	0,044	0,101	0,939	1,109
	1,499	2,702	3,014	3,202	3,538	3,766		
n.unc	0,023	0,007	0,041	0,119	0,111	0,046	2,047	1,733
	0,292	0,188	1,423	6,577	10,352	1,611		
n.unlp	-0,464	-0,308	-0,247	-0,131	-0,111	-0,266	1,821	3,174
	-6,357	-4,949	-7,619	-5,371	-6,509	-8,499		
n.unr	-0,440	-0,378	-0,264	-0,128	-0,101	-0,287	2,890	3,076
	-3,983	-7,527	-7,543	-6,175	-7,982	-9,809		
Adj -R Sq						0,102		
(b) Tn							39,331	20,920
Obs	3.816							

Note: t-values are given in the second line below each parameter estimate.
(a) THn: Test statistics testing whether each individual slope parameters satisfy the null hypothesis.
(b) Tn: Joint test statistic of the hypothesis that all the slope parameters of the model satisfy the hypothesis.

Source: Based on CEUN 1994

TABLE 2.3

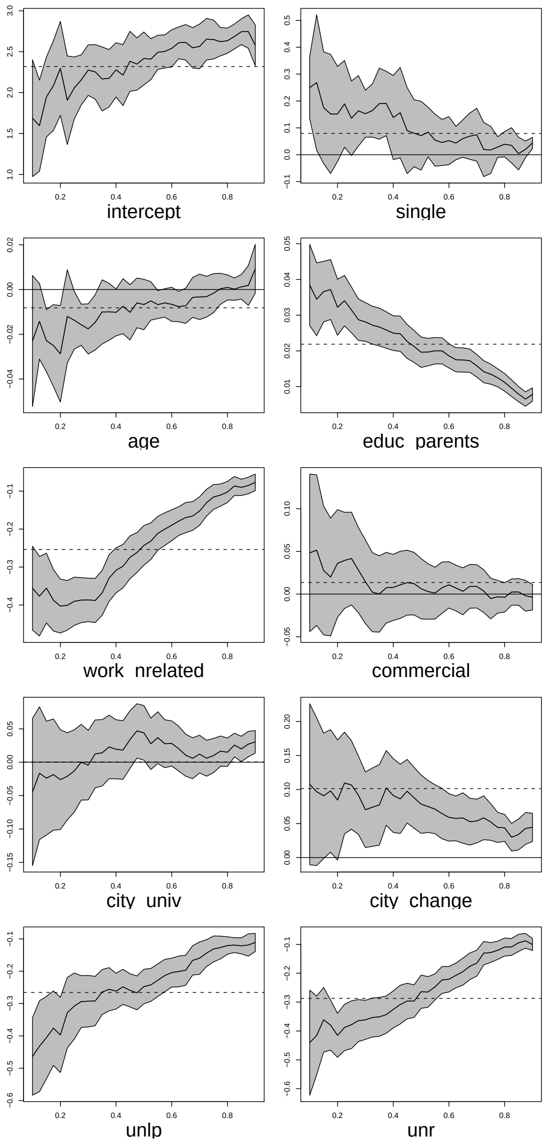
Quantile Regression Results. Law.

	Quantile					OLS	(a) THn. Null Hypothesis :	
Variable	0,10	0,25	0,50	0,75	0,90		Location	Location/Scale
(Intercept)	1,726	2,098	2,363	2,603	2,546	2,181	0,865	0,962
	8,675	15,000	24,638	36,780	38,895	27,077		
single	0,085	0,094	0,074	0,016	0,010	0,068	2,069	1,136
	0,887	1,605	2,090	0,589	0,408	2,107		
male	-0,060	-0,063	-0,052	-0,021	0,018	-0,027	1,732	2,479
	-1,353	-2,198	-2,982	-1,607	1,421	-1,605		
age	-0,010	-0,006	-0,002	0,000	0,006	-0,001	2,328	3,259
	-2,324	-1,753	-0,698	0,000	3,178	-0,294		
educparents	0,024	0,019	0,015	0,011	0,011	0,019	2,121	0,764
	4,283	5,415	7,079	7,441	7,792	9,084		
nworkrelated	-0,064	-0,032	-0,004	0,000	0,018	-0,019	0,822	0,839
	-1,181	-0,922	-0,201	0,000	1,247	-0,807		
nworknotrelated	-0,276	-0,280	-0,166	-0,089	-0,061	-0,186	1,915	2,458
	-5,229	-8,608	-8,424	-6,073	-4,672	-9,477		
commercial	-0,041	-0,006	0,006	-0,008	-0,006	-0,001	1,485	1,381
	-0,841	-0,217	0,338	-0,664	-0,506	-0,051		
private	0,127	0,076	0,038	0,027	0,036	0,058	0,882	1,953
	2,802	2,701	2,279	2,215	3,143	3,349		
cityuniv	0,040	0,005	0,025	0,011	0,021	0,004	0,800	0,927
	0,812	0,148	1,168	0,805	1,648	0,184		
citychange	0,058	0,072	0,052	0,041	0,053	0,067	1,060	0,914
	0,939	2,010	2,489	2,608	3,171	3,188		
n.unc	-1,250	-0,982	-0,752	-0,580	-0,482	-0,799	2,032	2,757
	-17,051	-21,916	-26,206	-26,937	-31,484	-35,656		
n.unlp	-0,624	-0,447	-0,282	-0,192	-0,128	-0,333	2,038	2,093
	-6,876	-7,122	-6,956	-8,154	-5,223	-11,002		
n.unr	-0,578	-0,376	-0,433	-0,373	-0,318	-0,418	0,871	1,564
	-5,963	-9,839	-20,664	-16,592	-13,945	-15,382		
Adj -R Sq						0,261		
(b) Tn							17,225	15,790
Obs	4812							

Note: t-values are given in in the second line below each parameter estimate.
(a) THn: Test statistics testing whether each individual slope parameters satisfy the null hypothesis.
(b) Tn: Joint test statistic of the hypothesis that all the slope parameters of the model satisfy the hypothesis.

Source: Based on CEUN 1994

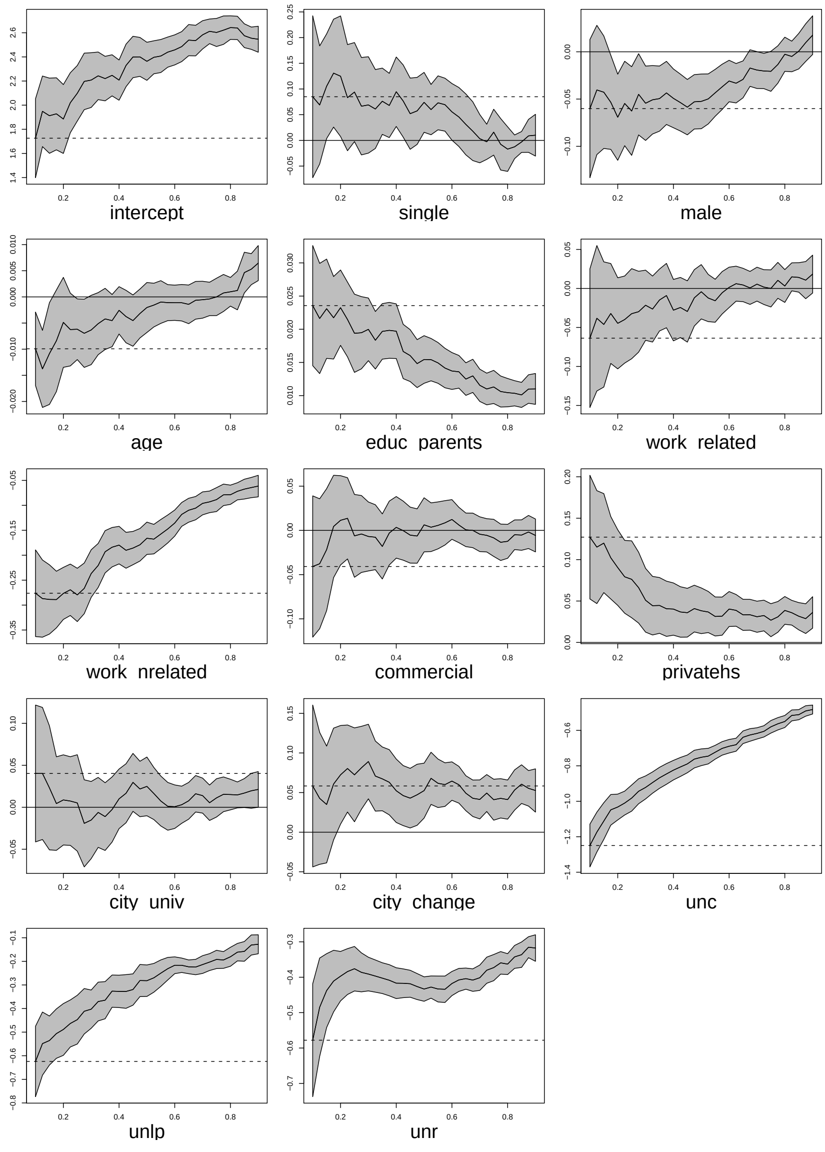
FIGURE 2.2
Quantile Regression Results: Accountancy



Source: Based on CEUN 1994

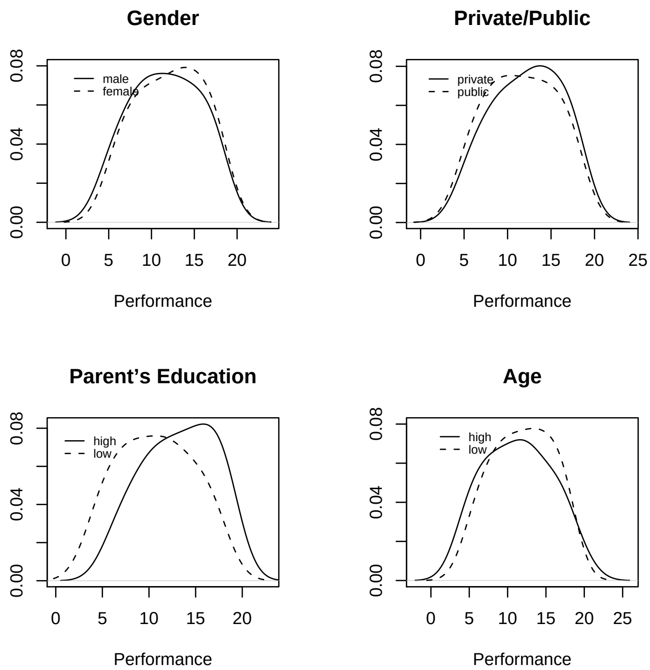
FIGURE 2.3

Quantile Regression Results: Law



Source: Based on CEUN 1994

FIGURE 2.4
Conditional Densities. Accountancy.



Source: Based on CEUN 1994

the shaded area represents a 90% confidence interval. The dotted horizontal line represents the OLS estimation. When relevant, the solid horizontal line simply indicates zero.

2.5.1 Effects of Individual Factors

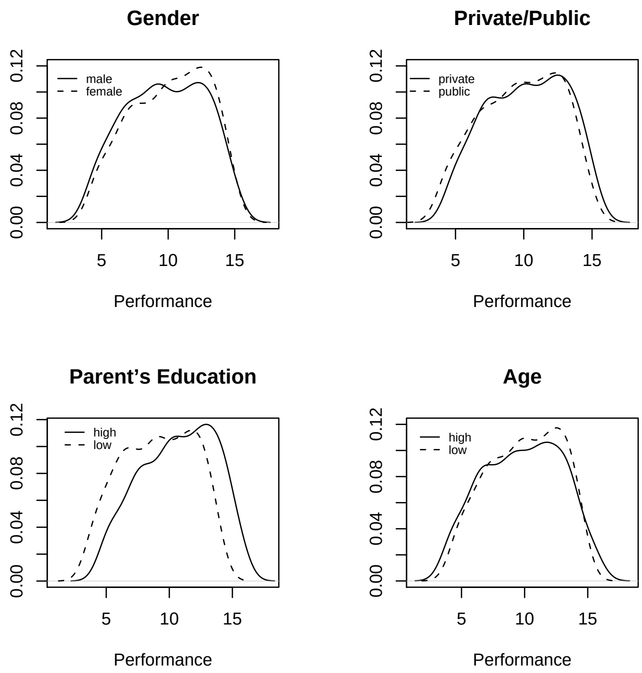
Now we turn to the analysis of the effect of individual factors. We will start by commenting on the results for Accountancy and then note differences and similarities with respect to Law. The gender dummy has a negative and significant effect in the mean OLS based model for Accountancy, suggesting that the expected performance of males is around 6% lower than that of females. Nevertheless, quantile regression results reveal that the effect is stronger in the lower levels of the conditional distributions, decreasing in absolute values and becoming statistically insignificant at the upper level.

Figure 2.4 illustrates this point by showing the conditional densities of performances of Accountancy students, for males and females, with all the remaining covariates set at their mean levels.[30] Even though the effect is in general mild, the estimated densities show that

[30]The Appendix describes the procedure used to estimate these densities.

FIGURE 2.5

Conditional Densities. Law.



Source: Based on CEUN 1994

the conditional distribution of males' performances has a larger left tail, so gender differences appear mostly in this range and not among those with higher performances.

The case of Law students, illustrated graphically in Figure 2.5, is slightly different, since the gender dummy is significant only in the middle part of the conditional distribution and insignificant in the extremes; consequently, the conditional distribution of performances for females is skewed to the right as compared to that of males. Though the effect is mild, it appears clearly in Figure 2.5.

An important issue is the effect of private versus public secondary education. The positive OLS based effect is actually a consequence of a positive and significant effect in the center of the conditional distribution of performances, in spite of being insignificant in the extremes. Public secondary schools in Argentina are, overall, perceived to be of lower quality than private ones since they usually receive students with less favorable socioeconomic background, with the exception of a few ones who are very traditional and manage to attract the very best students. Consequently, the fact that effects are nil in the extremes and positive in the center is compatible with the idea that once in college students from public secondary schools have a markedly negative asymmetric distribution of performances with most in the lowest tail of the distribution and relatively few at the top. The opposite results appears in the case of those with private education: most students have a good performance and relatively few of them have extremely bad performances. In either case the very good students, in terms of their performances, do not seem to have benefited from having attended one type of school or the other, and the same happens in the other extreme. This result is illustrated graphically in Figure 2.4, where the conditional densities of performances are plotted for students with private and public secondary education. The central part of the conditional distribution for those with private secondary school education appears shifted to the right, with the extremes unaltered, compared with those students with public secondary school background, compatible with positive effects in the middle and nil effects in the extreme. The case of Law students is different since private education has a strong effect in the bottom of the conditional distribution, decreasing monotonically and having a rather constant effect beyond the quantile 0.4.

Regarding parental education, as expected, the mean effect is positive, implying that students with a better parental background are expected to perform better. Quantile regression results provide relevant additional information suggesting that this effect is clearly heterogeneous and much stronger in the bottom of the distribution. A similar effect is found with Law students. This is consistent with a decreasing returns effect (semi-elasticities, in our case) where if we start at the bottom of the distribution of unobserved factors and measure the effect of increased family background, we should expect it to be positive but marginally decreasing as we move progressively towards groups of individuals more favored in their unobserved factors. Graphically, Figures 2.4 and 2.5 plot the conditional densities for individuals with parental with 7 and 18 years of education, respectively. The distribution of performances of students with more educated parents is shifted to the right and more skewed to the left, consistent with the effect of parental education being positive but decreasing across the quantiles. This is suggestive of substitutability between parental education and ability.

Age effects are interesting. OLS estimations are insignificant for both accountants and lawyers, suggesting that age has no effect on the conditional mean of performances. Nevertheless, the age effect by quantiles ranges monotonically from being significantly negative in the lower levels to slightly significant and *positive* in the upper quantiles; a very similar and stronger effect is also found for lawyers. This seems to be indicative of a pure *scale* effect where, all else being equal, classes with older students are more disperse in the sense that age plays a positive role for those in the upper tail of the distribution of non-observables and a negative one for those conditionally in the bottom. This is consistent with the intuition that good but otherwise older students may be more focused and mature about what they expect from their education (the positive effect of age) and hence perform better than those in the bottom (badly motivated or low skilled) for whom age plays a negative role in their performance. This result can be seen in Figures 2.4 and 2.5, where we plotted the conditional distribution of performances for individuals who, at the moment of the census were 21 and 30 years old. Consequently, in spite of having similar locations, the conditional distribution of performances of older students is more disperse than that of younger ones. As advanced in the Introduction, the insignificance of the age variable in the OLS mean model might lead careless observers to the erroneous conclusion that age has no effect in performance, ignoring the fact that it has a non-trivial effect on the dispersion, a fact that has important consequences since more heterogeneous groups may require a different pedagogical treatment to younger and more homogeneous ones.

Next we explore the effect of working while studying. Variables *work-related* and *work-not-related* are dummy variables both indicated with one if the student has a job related to her subject of study, and whether she works in an unrelated job, 'not work' being the implicit omitted category. OLS results suggest a negative effect on performances: overall, jobs affect performances negatively, with a stronger effect in the case of those working in jobs not related to their careers. Quantile regression provides a more accurate characterization. Consider first the case of accountants; once again, both effects are stronger in the bottom of the conditional distribution of performances. Interestingly, the effect of working in non-related jobs is consistently negative and significant, but the effect of working in related jobs does not have a significant effect above the median. These are very relevant results since they imply that career specific jobs do not compromise performance for Accountancy students with relative good performances[31]. The case of Law students is different. The dummy variable denoting jobs related

[31]The Internship Law (Ley de Pasantías - National Law 25165-99) is quite explicit regarding the complementary

to their career is never significant at all quantiles and in the OLS model. The effect of working in non-related jobs is similar to that for accountants. It is worth noting that, as stressed in the introduction, these effects cannot be interpreted as causal. The endogeneity of these variables has not been address. For instance, it could happen that because performance at University was low, students decided to enter into the labour market. With this caveat in mind, results seem to sugget that the dynamics of a career in Law is compatible with a job related to the practice of the Law without affecting performance, but also with the fact that students who do not work do not have a better performance than those who have jobs related to the career. The detrimental effect appears only in the case of those working in jobs not related to the career.

A much debated topic in the local literature is the relevance of the type of secondary education, where students who have the 'commercial' orientation are expected to have a relative advantage in Accountancy. Surprisingly the type of secondary school orientation has a homogeneous, not significant, effect on performances in both Accountancy and Law.

Marital status is homogeneously non-significant across most quantiles, for both Accountancy and Law students. Location variables have rather homogeneous effects, so quantile regression results do not add much to those revealed by OLS. Having to commute to attend college is not a relevant factor across all quantiles of the conditional distribution of performances. The fact that students relocate to attend college has a homogeneously relevant and positive effect in performances, much in accordance with the idea that those willing to pay the fixed costs of relocation are the relatively good students.

2.6 Conclusions

The main goal of this paper is to measure the effect of observable individual characteristics on the whole conditional distribution of performances. One of the main reasons for choosing this strategy is that in the case of educational policies it is necessary to complement the standard educational production function approach, by studying not only the mean effects of observable variables but also their impact on the shape of the distribution of performances. This is relevant since educational policies are often expected to promote equality of opportunities and possibil-

nature of internships, defining them as 'supervised practices related to specialization and training' (Art. 2) in order to obtain 'practical experience to complement theoretical training'.

ities, and hence distributive outcomes matter. Also, if policy actions are oriented towards the less advantaged, or any other specific group, it is important to assess whether the impact of a policy measure is homogeneous for all students, or whether average effects are actually an imprecise summary of a more complex reality that may systematically benefit certain individuals more than others.

Heterogeneities arise from interactions between unobserved and observed factors in the production of educational outcomes. Quantile regression methods are shown to provide a flexible framework to model these interactions between observed and unobserved factors, which are the source of non-homogeneous effects on performance that alter its conditional distribution in subtle ways improperly summarized by mean OLS based methods.

This methodological framework is adopted and applied to the case of college students in Argentina, whose social and institutional characteristics, that combine free access, a flexible schedule and a diverse socio-economic composition of students, provide ample sampling variability making it a relevant case study.

The empirical results of our research strongly suggest the presence of heterogeneous effects, which leaves ample room to question whether relevant factors such as parental education or secondary school type are stronger or weaker for certain individuals. The results of this paper indicate that, overall, effects are found to be less relevant in the top of the distribution, in the sense that all factors that contribute positively to performance (better family background, not having to work, etc.) are stronger in the bottom. Hence, policies that enhance the possibilities of students initially in the lower part of the distribution have the dual effect of increasing their absolute performances (through their positive effect) and reducing disparities due to their stronger effect in this group of students. These results are important for the design of educational policies aimed at promoting equal opportunities, as along the lines advanced by Roemer (1998), they must be tailored to compensate with external resources the different circumstances faced by students exerting similar levels of effort.

2.7 Appendix

2.7.1 Estimating Conditional Densities

In order to estimate the density of performances (y) conditional on a vector x of explanatory variables, we first obtain a random sample from the conditional distribution $y|x$. Machado and Mata (2005) suggest the following procedure to obtain random numbers based on an estimated model for the conditional quantiles. Assume that τ is a random variable uniformly distributed in (0,1). By the probability integral transformation theorem, if $y|\, x \sim F(Y|\, x)$

$$Q_\tau(y|\, x) = F^{-1}_{y|x}(\tau) = x'\beta(\tau) \sim F_{y|x}.$$

Then, we can obtain a random sample of size J of $y|\, x$ by first generating uniformly distributed random numbers $\tau_j, j = 1, \ldots, J$, and then computing $x'\hat{\beta}(\tau_j), j = 1, \ldots, J$, where $\hat{\beta}(\tau_j)$ are the estimates of the coefficients of the linear quantile regression for quantiles $\tau_j, j = 1, \ldots, J$. In our case, the vector x is set at convenient values. For example, in the comparison between students with public vs. private secondary education, two samples were obtained by setting x at their sample averages, and then switching the dummy variable for secondary school background from zero to one.

Once a random sample of $y|x$ is available, an estimate of the conditional density is obtained by applying standard kernel methods to it. The equivariance property of quantiles makes it straightforward to extend this mechanism to obtain random samples of any monotone transformation of y. In our case, since the model is estimated for the logs of performance, it is easy to see that $\exp(x'\beta(\tau_j)), j = 1, \ldots, J$ is a random sample of the original variable in levels, when the model is estimated in natural logarithms.

References

1. Araoz, A., 1968, Comentarios sobre el trabajo del Dr. Julio H.G. Olivera: La Universidad como Unidad de producción, *Económica*, La Plata, No 1.

2. Betts, J. R and D. Morell, 1999, The determinants of undergraduate grade point average. The relative importance of family background, resources and peer group effects, *Journal of Human Resources*, 34(2), 268-293.

3. Card, D. and A. B. Krueger, 1996, The economic return to school quality, in WW.E.Becker and W.J. Baumol (eds.), *Assessing Educational Practices: The contribution of Economics*, The MIT Press.

4. Chesher, A., 2003, Identification in nonseparable models, *Econometrica*, 71, 1401 - 1444.

5. Coleman, J.S., E. Cambell, C. Hobson, J. McPatarland, A. Mood, F. Weinfeld and R. York, 1966, *Equality of Educational Opportunity*, Washignton DC, US Government printing office.

6. Di Gresia L, M.V.Fazio, A. Porto, L Ripani and W. Sosa Escudero, 2007, Academic Performance of Public University Students in Argentina, *Well-being and Social Policy*, Vol, 3, No 2, second semester.

7. Ehrenberg R. G., 2004, Econometric studies of higher education, *Journal of Econometrics*, 121, 1-2, 19-37.

8. Eide E. and M. Showalter, 1998, The effect of school quality on student performance: a quantile regression approach, *Economics Letters*, 58, 345-360.

9. Figlio, D., 1999, Functional form and the estimated effects of school resources, *Economics of Education Review*, 18, 241-252.

10. Giovagnoli, P., 2005, Determinants of university desertion and graduation: an application using duration models, *Económica*, La Plata, 1-2, 57-88.

11. Gonzalez Rozada, M. and A. Menendez, 2002, Public university in Argentina: subsidizing the rich?, *Economics of Education Review*, 21, 341-351.

12. Hanushek, E., 1979, Conceptual and empirical issues in the estimation of education production functions, *Journal of Human Resources*, 14(3), 351-388.

13. Hanushek, E., 1986, The economics of schooling: production and efficiency in public schools, *Journal of Economic Literature*, 24(3), 1141-1177.

14. Koenker, R. and Bassett, G., 1978, Regression quantiles, *Econometrica*, 46(1), 33-50.

15. Koenker, R., 2005, *Quantile Regression*, Cambridge University Press, Cambridge.

16. Koenker, R. and Xiao, Z., 2002, Inference on the quantile regression process, *Econometrica*, 70(4), 1583-1612.

17. Levin, J., 2001, For whom the reductions count? A quantile regression analysis of class size and peer effect on scholastic achievement, *Empirical Economics*, 26, 211-246.

18. Machado, J. and Mata, J., 2005, Counterfactual decomposition of changes in wage distributions using quantile regression, *Journal of Applied Econometrics*, 20(4), 445-465.

19. Mas-Colell A., M.D.Whinston and J.R.Green, 1995, *Microeconomic Theory*, Oxford University Press.

20. McGuckin, R. and D. Winkler, 1979, University resources in the production of education. *Review of Economics and Statistics*, 61 (2), 242-248.

21. Naylor R. and J. Smith, 2004, Determinants of educational success in higher education. In Geraint Johnes and Jill Johnes (eds.), *International Handbook in the Economics of Education*, Edward Elgar, Northhamp-ton, MA.

22. Olivera, J.H.G., 1964, Aspectos Económicos de la Educación, Reprinted in *Economía Clásica Actual*, Editorial Macchi, Buenos Aires, 1977.

23. Olivera, J.H.G., 1967, Die Universitat als Produktionseinheit, Weltwirtschaftliches, Band 98, Heft 1.

24. Porto, A. (ed.), 2007, *Mecanismos de admisión y rendimiento académico de los estudiantes universitarios. Estudio comparativo para estudiantes de Ciencias Económicas.* Edulp, La Plata, Argentina.

25. Roemer, J. E., 1998, *Equality of Opportunity*, Harvard University Press, Cambridge.

3 The Role of the Labour Market on Youth Schooling Decisions

3.1 Introduction

The role of labour market conditions on youth schooling behaviour has been empirically assessed in different countries using a wide range of data and methods. It can be framed within Becker´s model, in which students decide every year to leave school based on the expected net benefits of an additional year of schooling. That is, when labour markets conditions worsen, all else equal, the opportunity cost of attending school decreases, raising the expected net benefits of schooling and lowering dropout rates.

In Argentina, studies are mostly related to youth schooling behaviour with household socio-economic conditions and school characteristics, giving marginal attention to the influences of labour market conditions, see for instance Marchionni and Sosa Escudero (2000), Binstock and Cerrutti (2005), Paz (2008).

In this chapter, I empirically examine whether changes on labour demand affect the probability of secondary school dropout using a novel dataset from Argentina - *Educacion y Empleo de los Jovenes* -EJJ hereafter - that allows a tracing of individual education histories from early in life in a retrospective manner.

The primary advantage of this dataset is that information on the year of entrance at school and the year of exit from school are both available, providing me with a unique opportunity to model dropping out as a dynamic event in which the probability of leaving school can vary depending on how many years a person has already spent in school. In addition, most school characteristics and socio-economic conditions of individuals at the time that schooling choices were made are known, offering a large set of controls.

The EJJ was implemented in 2005 to young people between 15 and 30 years old living in the main urban area of Argentina, Greater Buenos Aires[32]. The fact that these cohorts were exposed to different labour market conditions by the time they reached secondary school is exploited in this study to identify changes in labour demand.

Using data from EPH surveys, I construct proxies for youth labour demand - i.e youth unemployment rates, by gender, level of education and year. These variables are matched to the EJJ sample to estimate the effect of labour demand on dropout rates by applying a discrete time proportional hazard model.

The organisation of this paper is as follow. The next section motives the analysis by pre-

[32]This area constitutes 30% of the total number of students in the country.

senting trends on dropout rates and labour demand for the cohorts under study. Section 3 is divided into two parts; the first part is devoted to the description of data used in this study and how it has been collected, while the second part revises and justifies the empirical approach. Section 4 discusses the estimation results obtained and the final section concludes.

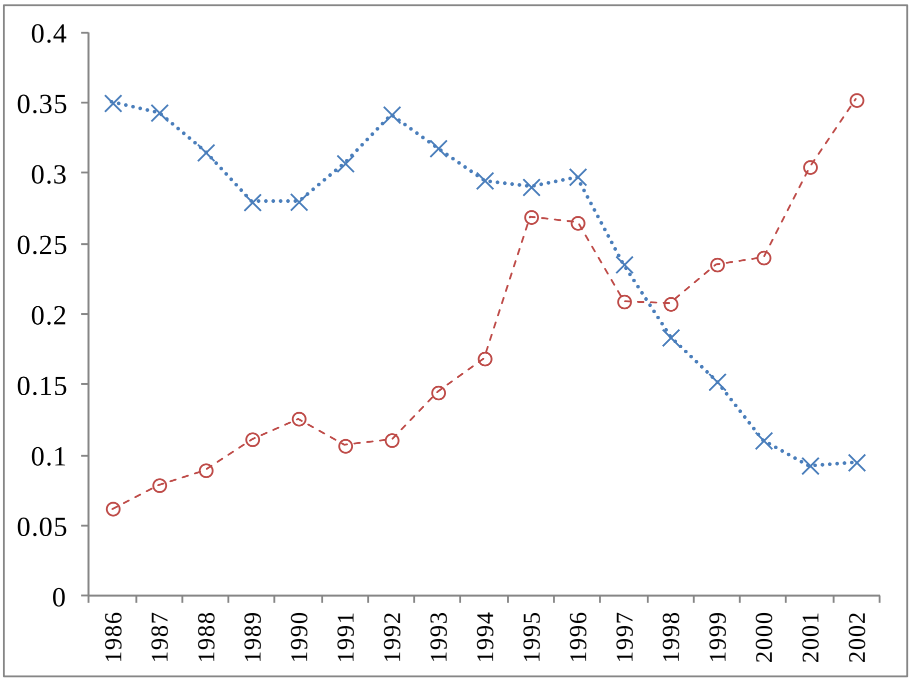
3.2 Motivation

Since it is one of the aims of this chapter to shed light on the role of labour market conditions on youth schooling decisions using the EJJ sample, it is important to look in some detail at the youth unemployment rates that these cohorts experience by the time they reached secondary school age. When youth unemployment is low - and a high probability of obtaining a job is anticipated - the opportunity cost of studying increases and some young people may simply decide to drop out of school and enter into the labour market. That is, as the human capital theory suggests (Becker, 1964), all else equal, the costs will exceed the expected future benefits from an additional year of education, i.e. the net present value of investing in education will be negative.

Figures 3.1 and 3.2 depict youth unemployment rates and dropout rates for young males and females between 1987 and 2002 respectively. Note that by 1988, people born in 1975 were 13 years old and were 30 years old by 2005 - when the survey took place. These estimates were calculated using cross-section micro-data from urban household surveys -*Encuesta Permanente de Hogares*, EPH - for Greater Buenos Aires area. Dropout rates refer to the proportion of individuals aged 13-17 who reported not attending school in the survey year. Youth unemployment rates are calculated for individuals aged 18-24. Similar patterns to the ones presented below were encountered using narrowed age group definitions.

FIGURE 3.1

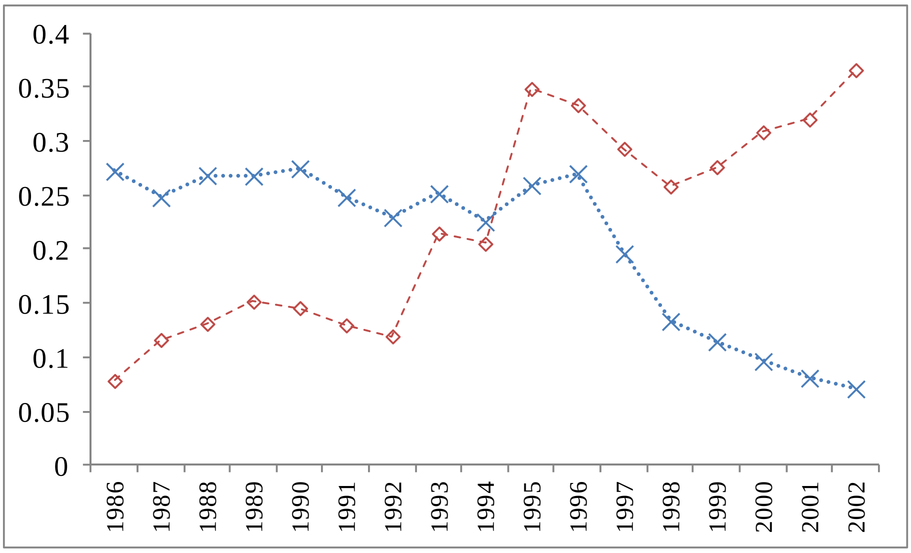
Youth Unemployment and Dropout rates. Males.



Reference: -x-x- Refers to Dropout Rates. –o–o–o Refers to Youth (Male) Unemployment Rates

Source: EPH 1987-2002. Geater Buenos Aires.

FIGURE 3.2
Youth Unemployment and Dropout rates. Females.



Reference: -x-x- Refers to Dropout Rates. –o–o–o Refers to Youth (Female) Unemployment Rates

Source: EPH 1987-2002. Geater Buenos Aires.

Focusing on males, some interesting patters emerged.

While in 1985 35 percent of young people were out of the education system, only 10 percent did not attend secondary school 15 years later. These figures are consistent with Population Census data. This data reveals that 39 percent of people aged 15 in 1980 were out of school, while data for 2001 shows that the non-attendance rate decreased to 14 percent.

Interestingly, the youth unemployment rate seems to move in the opposite direction compared to the dropout rate. On the one hand, during the late 1980s, the rise in youth unemployment has been accompanied by a decline in the dropout rate, suggesting that worse labour conditions encourage young people to stay at school longer. On the other hand, between 1991 and 1994, youth unemployment fell slightly or at least did not rise, while the proportion of males not attending school registered a peak.

By the end of 1994, when the Tequila crisis hit the country, unemployment jumped to unprecedented levels. The role of unemployment in schooling decisions is less clear for 1995 and 1996, years for which no correspondence between unemployment and dropout is observed. This situation could be the result of a strong contraction in the economy, thus negatively affecting household income and consumption.

When turning attention to females, unemployment and schooling variables appear to behave slightly differently compared to those of males. While female dropout rates do show a similar decreasing trend across the whole period as in the case of males, dropout rates seem to move in the same direction of unemployment, at least until the late 1990's.

Another interesting aspect of this graph is that females are considerably less likely to be out of school, with non-attendance rates never higher than 27 percent.

The analysis suggests that during periods of high labour demand, males, rather than females, tend to leave school as they may perceive an increasing opportunity cost of studying versus working.

3.3 Data and Empirical Approach

3.3.1 The EJJ

The data used in this chapter comes from a specific questionnaire known as EJJ applied in 2005 to the eligible population - young people aged 15 to 30 living in Greater Buenos Aires.

The questionnaire was carefully designed with the aim of capturing educational paths as well as labour market experiences. It was applied as part of the official urban household survey - EPH.

The sample consisted of dwellings already visited in previous waves of the EPH in which at least one eligible population was living there.

The importance of this data is highlighted by the rich set of variables containing retrospective information on the calendar years when the person started and ended primary and secondary school as well as grades completed and grades repeated (if any). As is explained in detail below, with this information, it was possible to re-construct the educational history for each individual and to measure the length of enrolment in secondary school until graduation (or dropout), and any indication of delay due to failure.

The degree of accuracy that can be obtained from retrospective data has been widely discussed in the literature. One obvious problem is that the older the interviewee, the less likely he/she is to remember correctly whether and when certain events happened in the past. There is, however, an extensive body of research suggesting how the recall bias can be reduced (see Mathiowetz et.al., 2001, Eisenhower, et al 2004). The application of specific techniques - such as landmark events - facilitates the reconstruction of the past. Furthermore, it has been shown that certain events, like marriage, births and schooling can be collected within acceptable levels of precision (Dex, 1991).

The EEJ adopted these techniques achieving very high levels of individual response. Indeed, out of 880 selected individuals, 807 of them answered the questionnaire; see INDEC, Modulo sobre Educacion y Empleo de los Jovenes (2005) for more details.

The questionnaire also collects information on many school characteristics, socio-economic and personal conditions at the time that schooling choices were made, in turn providing a large control set of observable variables. It encompasses basic features of primary and secondary school, such as whether the person attended a public or private (religious or not religious) school; whether he/she did not change school during secondary education; whether schools were located in Conurbano or in the City of Buenos Aires; whether the person attended a vocational education or whether the person attended extended education. In Argentina, people can attend school both in the morning and in the afternoon (extended education) or just in the morning or the afternoon (simple education or half time education). The individuals also reported whether they had textbooks available during their studies and their parents' education even when they were not living with their parents.

3.3.2 Empirical Approach

Taking advantage of the detailed information on educational histories available in the EJJ data, I will examine the school leaving behaviour as an event that could happen at any time while the student is enrolled at secondary school using duration models, also known as transition models, Cox (1972); Lancaster (1979); Kalbfleisch and Prentice (1980); Cox and Oakes (1984) and Klein and Moeschberger (1997).

Unlike the traditional -static- model in which the probability of school dropout is examined at a point in time, with the in duration model it is possible to analyze whether the risk of leaving school varies over time by computing the conditional probability that dropout will occur given that it has not yet occurred.

The depend variable of interest is the time spend in school, T. T is defined as a discrete, non-negative random variable, representing the time enrolled at secondary school and is measured in years. The year refers not to chronological year but to the year the student is enrolled in at secondary school. In the sample under study, the time spans from the 1st year until the 7th year - 7 is the maximum amount of years a student reported of staying in secondary school[33]. For

[33] Although the curriculum mandates 5 years of study to obtain a diploma, some people take longer

instance, the 1st year is 1988 for those who first enrolled at secondary school in 1988, the 1st year is 1989 for those who first enrolled at secondary school in 1989 and so on. In this way, the time of origin is the same for all individuals.

The duration model is also applied to analyse how different covariates amplify or dampen the risk of leaving secondary school, with a particular focus on understanding the role of youth labour market conditions.

There are some methodological considerations to take into account before applying duration analysis to the EJJ data.

First, although the decision of leaving school can be conceptualised as being fundamentally continuous, the time can only be measured in discrete units - year intervals.

Second, the discrete nature of data results in "tied data"- more than one student experiences the event within the same interval of time in my dataset.

Third, data is right censoring, and censoring occurs for two reasons: (i) because the person is still in secondary school by the time of the interview, and (ii) because the person left school due to a different (competing) cause, such as graduation. In this analysis, I assume independent censoring, which means that each year individuals in the risk set do not differ systematically from censored individuals, Singer and Willet (1991).

The approach that deals with these methodological considerations and properly accounts for discrete hazard functions is the one proposed by Prentice and Gloeckler (1978). I adopt this approach - a discrete time reduced form hazard specification - to analyze the data.

$\lambda(X_{it},(t)) = 1 - \exp(-exp(X_{it}'\beta + \alpha_(t)))$

where $\lambda(X_{it},(t))$ is the hazard rate in discrete time for person i at interval of duration of t, X_{it} is a vector of covariates summarizing observed differences between individuals at t, β is a vector of parameters to be estimated. It is a proportional hazard model like the continuous Cox model. The main difference is the interpretation of the hazard function. In the discrete case, it is the conditional probability, while the continuous case makes reference to the instantaneous hazard rate. $(\alpha_(t))$ are dummy variables associated with the different times- the baseline hazard function . An advantage over other approaches is that this underlying baseline hazard function is non-parametric estimated. The incorporation of dummy variables associated with the different times provides a direct estimate of the baseline hazard function and prevents inconsistency estimation of the coefficients due to missespecification of the baseline hazard.

An additional advantage is that it can be estimated using complementary log-log regression $log(-log(1-\lambda_i)) = (X_{it}'\beta + \alpha_(t))$, applied to re-structured data as explained below.

In Section 3.4, I also show the results based on the model proposed by Bruce Meyer (1990, 1995). The model is a modification to the PG model. It allows for unobserved heterogeneity by incorporating a random variable with Gamma distribution. This random variable summarises the impact of a group of covariates that affect the risk of the event but are not observed in a direct manner. This is because they are intrinsically unobservable, or because the data is unavailable. Other interpretations include possible mistakes in the measurement of data. Lancaster, (1990).

Before analyzing the results applied to the EJJ data, the next subsection describes the final sample used in the analysis and the limitations encountered in the data.

3.3.3 The Sample

For the purpose of this analysis, certain group of individuals were excluded from the sample due to different reasons.

Out of 807 interviewed, 69 were excluded for not belonging to the risk group - never started secondary school, another 13 persons were attending adult education at the time of the interview as they were older than 21 years old and 2 other persons were in special education (i.e. persons with mental or physical problems). I also excluded 26 persons with inconsistencies between the years of birth and the year of starting secondary school. The inconsistencies could be either because the person misreported or because during digitalisation, the data was wrongly entered. Additionally, 12 persons were deleted due to missing values in some of the variables referring time and 55 young people who belong to the youngest cohort (those who started schooling in 2003 or 2004) to avoid severe censure.

The final sample contains a total of 631 observations; 358 finished secondary school and got a degree, 131 dropout school and 142 are still attending.

Due to the nature of data collection it is possible that some sort of selection bias exists. The older cohorts might not represent the true group of people starting school at that time. This is because migration to and from other places is more likely to occur in adult populations compared to young populations. And if migration occurred as a response to the labour market, the situation could be problematic. In order to have an idea about the magnitude of this problem, I examined two additional questions collected within the EPH survey. These two questions provide information on: (a) the place where the person lived 5 years ago and (b) the locality the person was born. Looking at my sample, 95.6 percent reported that they were living in Greater

Buenos Aires five years ago and 87 percent reported being born in that locality. These figures suggest that if there is any selection bias, it would not invalidate the analysis.

Summary Statistics

Variable means for the total sample of 631 individuals are reported in the first column of Table 3.1.

Of the 631, 46 percent are males, 10 percent have parents with incomplete primary education and 48 percent have parents with primary complete or incomplete secondary education. The average age at entry into secondary school is 13 years old.

TABLE 3.1
Summary Statistics

	All		Dropout	
	Mean	SD	Mean	SD
Male	0.46	0.50	0.53	0.50
Incomp. primary or less (parents)	0.10	0.30	0.21	0.41
Completed primary or incomp.secondary (parents)	0.42	0.49	0.57	0.50
Completed secondary or more (parents)	0.48	0.50	0.22	0.42
Age first time enrolled in secondary educ.	13.08	0.85	13.51	1.22
Repeat at least once in primary	0.07	0.26	0.14	0.35
Repeat at least once in secondary	0.25	0.44	0.47	0.50
School located in Conurbano	0.71	0.46	0.78	0.42
Half-time education	0.78	0.41	0.83	0.38
Never changed sec. school	0.76	0.43	0.80	0.40
Bilingual education	0.16	0.37	0.09	0.29
Private (religious) educ.	0.19	0.40	0.06	0.24
Private (non-religious) educ,	0.15	0.36	0.07	0.25
Public educ.	0.66	0.48	0.87	0.34
Vocational education	0.13	0.33	0.13	0.34
Books available at home	0.83	0.37	0.70	0.46
Total obs	631		131	

Source: EJJ. 2005

While 7 percent of the sample reported had repeated at least one year during primary school, 25 percent repeated at least one year during secondary school. Around 70 percent of people in the sample reported they had attended a school located in Conurbano, probably because they lived in that area. The poverty rate in Conurbano has been historically higher compared to Ciudad de Buenos Aires. A high proportion of young people attend half time education (compared to full-time education). Around 24 percent of pupils changed secondary school at least once during their educational career. Attending a private school is less likely than attending a public school. Moreover, only 13 percent of people interviewed attended a vocational education school. Finally, 83 percent of the sample reported they had textbooks at home for studying at secondary school.

Next, the means for the total sample in column (1) are compared to the mean for those who left secondary school without getting a degree in column (3). On average, leavers are more likely to be males, have parents with very low education, are slightly older by the time they first enrolled in secondary school, have double the chance of being repeated (either primary or secondary), and are more likely to attend a public school and have less textbooks at home.

All of the variables already described come from the EJJ questionnaire. Table 3.2 presents summary statistics for two important variables to be included in the regression analysis: youth unskilled unemployment rates. These variables will act as a proxy for the labour market conditions faced by young people. I expect they capture the effect of the labour market on school decisions. The upper panel refers to the whole sample, while the lower panel shows the unemployment rates calculated for the subsample of dropouts. Interestingly, those who left school faced, on average, lower rates of unemployment compared to the whole sample[34].

TABLE 3.2
Youth, Unskilled Unemployment Rates

	All			
	Mean	SD	Min	Max
Youth (16-18), unskilled unemployment rate	0.423	0.154	0.139	0.744
Youth (18-21), unskilled unemployment rate	0.325	0.124	0.120	0.517
	Dropouts			
	Mean	SD	Min	Max
Youth (16-18), unskilled unemployment rate	0.387	0.159	0.139	0.744
Youth (18-21), unskilled unemployment rate	0.302	0.125	0.120	0.517

Source: EPH. 1987-2002

I constructed youth unskilled unemployment rates by sex, level of education and by year using the EPH for Greater Buenos Aires. Unskilled means without a secondary school degree. I considered two definitions for "youth": individuals aged 16-18 and individuals aged 18-24. These youth unskilled unemployment rates are then matched with EJJ files by sex and by calendar year when the individual in the sample reached 14 years old - this is the age that a person, by law, can start working.

[34] Average rates are slightly higher than those presented in Figures 1 and 2. This is because in this case the unemployment rate is estimated for those without a secondary degree.

3.4 Results

3.4.1 Estimates for Males

Table 3.3 presents the results for the discrete hazard model proposed by Prentice and Gloecker (1978) for males. The model does not include a constant for convenience presentation of the baseline hazard and is estimated by the method of maximum likelihood in a person-period dataset in which each row corresponds to a person- year observation[35]. The depend variable for each person-year is equal to 1 if the person dropped out of school in that year and is equal to zero otherwise. That is, each person contributes as many years as he/she was at risk of dropping out. The final dataset is an unbalanced panel. Youths who were enrolled in only the first year contributed 1 person-year each, while those enrolled for 4 years contributed 4 person-years. Those who, by the time of the survey were still enrolled, are considered right-censored and contributed the maximum of person-years observed. Similarly, if the individual was enrolled for 5 years and graduated that year, she/he contributed 5 person-years and is also treated as censored after year 5.

The signs of the coefficients indicate how a covariate affects the hazard rate. While a positive coefficient increases the hazard rate, a negative coefficient decreases the risk of dropping out.

As shown in column (1), all coefficients have the expected sign. Regarding parental schooling, the lesser the education level of the parents, the greater the risk of leaving school without completion. A student with parents who have only complete primary or incomplete secondary education has a risk of dropping out that is 2.2 times greater than a student in the base category (with at least complete secondary education). Parental education is a variable that is most often related to completion of secondary school, see Petrongolo and San Segundo (2002).

The age at entry into secondary school is also an important factor in explaining differences on the risk of abandoning school. Those who were older for their grade when started secondary school dropped out at a higher rate. If a student started secondary school with one year of overage, his/her risk of leaving school is 36 percent higher compared to a student who enrolled in secondary with the right age for his/her grade. [36]The most common factor leading to overage is repetition in primary school. In fact, the coefficient indicating whether a student repeats at least once in primary school is not different from zero. The lack of statistical significance is due

[35]The cloglog command in STATA was used in the estimations, (see Jenkins, 1997)

[36]By exponentiating the coefficient, the hazard ratio is obtained to calculate the factor change or percentage change in the baseline hazard associated with a one unit increase in a covariate.

TABLE 3.3

Estimates for the Discrete Time Proportional Hazard Model. Males

	(1)	(2)	(3)	(4)
Incomp. primary or less (parents)	0.426	0.457	0.408	0.441
	(0.454)	(0.455)	(0.452)	(0.453)
Completed primary or incomp. secondary (parents)	0.789**	0.808***	0.823***	0.845***
	(0.308)	(0.310)	(0.305)	(0.306)
Age first time enrolled in secondary educ.	0.328***	0.339***	0.359***	0.369***
	(0.116)	(0.117)	(0.121)	(0.121)
Repeat at least once in primary	0.421	0.386	0.423	0.391
	(0.379)	(0.383)	(0.382)	(0.386)
School located in Conurbano	0.874***	0.869***	0.855***	0.850**
	(0.331)	(0.330)	(0.331)	(0.331)
Half-time education	0.0104	0.0261	-0.0359	-0.0206
	(0.359)	(0.359)	(0.364)	(0.364)
Never changed sec. school	0.178	0.171	0.117	0.108
	(0.317)	(0.317)	(0.325)	(0.326)
Bilingual education	-0.761	-0.778	-0.679	-0.696
	(0.630)	(0.630)	(0.632)	(0.632)
Private (religious) educ.	-1.303**	-1.318**	-1.308**	-1.322**
	(0.610)	(0.611)	(0.613)	(0.613)
Private (non-religious) educ,	-0.257	-0.257	-0.248	-0.247
	(0.572)	(0.572)	(0.571)	(0.571)
Vocational education	-0.343	-0.329	-0.400	-0.385
	(0.396)	(0.396)	(0.401)	(0.401)
Books available at home	-0.222	-0.214	-0.198	-0.190
	(0.294)	(0.294)	(0.294)	(0.294)
Youth (16-18), unskilled unemployment rate	-3.789**	-3.787**		
	(1.597)	(1.592)		
Youth (18-21), unskilled unemployment rate			-3.644*	-3.664*
			(2.059)	(2.055)
Failure in Year t		-0.325		-0.328
		(0.486)		(0.485)
Baby boom	-1.084***	-1.077***	-0.969**	-0.966**
	(0.375)	(0.373)	(0.389)	(0.388)
Baseline Hazard (BH)				
Year 1	-7.300***	-7.464***	-7.931***	-8.076***
	(1.653)	(1.676)	(1.618)	(1.641)
Year 2	-6.750***	-6.865***	-7.384***	-7.481***
	(1.631)	(1.648)	(1.594)	(1.612)
Year 3	-6.357***	-6.468***	-7.001***	-7.092***
	(1.617)	(1.626)	(1.576)	(1.586)
Year 4	-6.246***	-6.385***	-6.870***	-6.991***
	(1.616)	(1.632)	(1.576)	(1.593)
Year 5	-6.044***	-6.180***	-6.664***	-6.779***
	(1.596)	(1.614)	(1.560)	(1.577)
Year 6	-5.581***	-5.733***	-6.250***	-6.383***
	(1.632)	(1.650)	(1.592)	(1.610)
Year 7	-5.131***	-5.299***	-5.792***	-5.943***
	(1.670)	(1.690)	(1.634)	(1.654)
Ho: BH is constant (Chi2(7))	28.41	28.43	31.85	31.76
Prob > chi2 =	0.00	0.00	0.00	0.00
Observations	1,223	1,223	1,223	1,223
DF	21	22	21	22
Chi2	428.1	427.5	432.5	431.9

Note: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Source: Author's calcluation using EJJ data.

to a high correlation with age at entry. When I dropped age at entry from the model, repetition at primary school turned significant at 1percent and none of the other coefficients changed; neither their sign nor their significance.

Students who attended schools located in Conurbano have a greater hazard rate compared to those attending schools located in the city, being 2.3 times. As explained earlier, living in Conourbano reflects poorer socio-economic conditions compared to those who lived in the City.

In terms of variables related to type of education, attaining a private (religious) education compared to public education significantly decreases the risk of dropout. Attending bilingual education or vocational education reduces the risk of dropout, although none of the coefficients are statistically significant from zero.

The unemployment rate has a negative and significant effect on the hazard rate of dropping out of secondary school. Estimates in columns (1) and (2) use as a proxy for labour market conditions the youth unskilled unemployment rate calculated for males aged 16-18, while estimates in columns (3) and (4) use the unemployment rates measured for the group of unskilled males aged 18-21. In both cases, unskilled means without a secondary school degree (incomplete secondary) and unemployment refers to the year when the student is aged 14. The results indicate that an increase of 5 percent in the youth unemployment rate would lead to a 16-17 percent reduction change in the hazard of leaving dropout rates depending on the measure of youth unemployment used[37].

In all these models I controlled for the demographic phenomenon occurred during 1975-1980 in Argentina. During these years, the Global Fertility Rate reached 3.4 kids per 100 women (See Figure 1 in the Appendix). In order to capture this phenomenon, I included a dummy equal to 1 for those people who belong to this group. Similar results - not reported - were obtained by replacing the dummy for specific demographic indicators such as the Global Fertility Rate.

It is worth noting that restricting the sample to include only those individuals born since 1980 - i.e excluding those cohorts that belong to the demographic explosion - all conclusions remain unchanged. The results are shown in Table 3.4.

In fact, the effect of the labour market it seems to be slightly bigger when these cohorts are

[37]Note that 2.5 percent of the sample enrolled in secondary school by the age of 15 and another 2.5 percent started between 16 and 18. For all these cases, unemployment is measured at the year they started secondary school. I rerun these regressions using adult unemployment in addition to youth unemployment (Petrongolo and San Segundo, 2002). However, the models were very unstable due to high multicollinearity problems.

TABLE 3.4

Estimates for the Discrete Time Proportional Hazard Model. Males. Restricted Sample

	(1)	(2)	(3)	(4)
Incomp. primary or less (parents)	0.0920	0.0765	0.0957	0.0731
	(0.606)	(0.607)	(0.608)	(0.611)
Completed primary or incomp. secondary (parents)	0.702*	0.698*	0.753*	0.746*
	(0.392)	(0.392)	(0.386)	(0.387)
Age first time enrolled in secondary educ.	0.461***	0.437**	0.514***	0.495***
	(0.168)	(0.171)	(0.170)	(0.172)
Repeat at least once in primary	1.085**	1.151**	1.085**	1.140**
	(0.488)	(0.497)	(0.485)	(0.492)
School located in Conurbano	1.452***	1.438***	1.463***	1.452***
	(0.549)	(0.550)	(0.558)	(0.560)
Half-time education	0.351	0.316	0.305	0.271
	(0.452)	(0.457)	(0.456)	(0.459)
Never changed sec. school	-0.281	-0.272	-0.332	-0.329
	(0.372)	(0.373)	(0.395)	(0.395)
Bilingual education	-0.424	-0.403	-0.322	-0.296
	(0.782)	(0.783)	(0.778)	(0.780)
Private (religious) educ.	-1.335*	-1.308*	-1.261*	-1.238
	(0.757)	(0.759)	(0.762)	(0.764)
Private (non-religious) educ,	-0.246	-0.237	-0.231	-0.230
	(0.638)	(0.637)	(0.628)	(0.627)
Vocational education	-0.572	-0.596	-0.624	-0.651
	(0.563)	(0.567)	(0.569)	(0.574)
Books available at home	-0.686*	-0.695*	-0.700*	-0.712*
	(0.376)	(0.378)	(0.379)	(0.381)
Youth (16-18), unskilled unemployment rate	-4.119**	-4.179**		
	(2.032)	(2.039)		
Youth (18-21), unskilled unemployment rate			-2.954	-3.026
			(2.640)	(2.643)
Failure in Year t		0.357		0.336
		(0.532)		(0.527)
Baseline Hazard (BH)				
Year 1	-9.164***	-8.789***	-10.45***	-10.12***
	(2.559)	(2.604)	(2.507)	(2.536)
Year 2	-8.390***	-8.099***	-9.675***	-9.421***
	(2.514)	(2.533)	(2.459)	(2.466)
Year 3	-8.229***	-7.913***	-9.547***	-9.277***
	(2.507)	(2.541)	(2.447)	(2.464)
Year 4	-7.944***	-7.606***	-9.238***	-8.946***
	(2.507)	(2.550)	(2.451)	(2.478)
Year 5	-7.845***	-7.502***	-9.101***	-8.810***
	(2.474)	(2.514)	(2.418)	(2.439)
Year 6	-7.414***	-7.043***	-8.739***	-8.417***
	(2.576)	(2.625)	(2.502)	(2.534)
Year 7	-4.391*	-4.042	-5.742**	-5.435**
	(2.481)	(2.529)	(2.403)	(2.439)
Ho: BH is constant (Chi2(7))	31.47	31.22	33.38	33.16
Prob > chi2 =	0.00	0.00	0.00	0.00
Observations	733	733	733	733
DF	20	21	20	21
Chi2	233.8	233.2	237.0	236.4

Note: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Source: Author's calcluation using EJJ data.

not included.

Finally, in columns (2) and (4) of Table 3.3, I included a time-varying covariate: a dummy equal to 1 if the student repeats a grade in year t and zero otherwise. However, it should be noted that failure during secondary school is a bad control (See Angrist and Pischke (2008)). The youth unemployment rate is likely to affect both repetition as well as the school leaving decision. On the one hand, prospective labour market conditions could make students miss school days which may lead them to repeat a grade. On the other hand, if unskilled youth unemployment rate is high, students would stay at school because they do not expect to find a job outside putting minimum effort on performance and therefore repeating the grade (Mirriless, 1981). The coefficient estimated is not significant in any of the specifications.

Except for the variable "failure", all the other variables included in the analysis so far do not vary with time. Furthermore, throughout the analysis, it has been assumed that the effect of a change in the explanatory variable over the hazard was independent from the interval of time in which it was measured. This assumption is relatively easy to check including interactions between covariates and year dummies (baseline hazard) and then testing the null hypothesis that the coefficients do not vary over time. The small sample size prevents me to do this analysis. However, it is important to point out that this assumption is much less restrictive in the discrete model than in the continuous model, in which proportionality had to take place “at each moment in time.”

Other interactions of interests could also be explored. For instance, one could test whether those who attended public school or a school in Conurbano were more sensitive to the impact of labour market conditions. Unfortunately, given the data at hand, it was not possible to explore these interactions as there were too little observations in the categories to estimate meaningful coefficients.

The final set of coefficients at the end of Table 3.3 refers to dummy variables corresponding to each year of enrolment. In this way, the baseline hazard is estimated non-parametrically without imposing a potentially inappropriate shape. Using a Wald test of equality of coefficients, I can reject the hypothesis null that the baseline hazard is constant – i.e. does not vary with time. The estimated coefficients for these dummies suggest instead that the hazard does rise over time.

3.4.2 Estimates for Females

Table 3.5 shows the results for females, using the same model applied for males.

Although parental education and age at entry are again very significant with the expected signs as in the case of males, there are some differences.

First, those who never changed school appear to have a significant higher risk of dropping out compared to those who did change at least once in secondary school. One explanation for this result could be that changing school acts as a proxy to parental attitudes towards education. For instance, if the student is not performing adequately in school, parents who care about the education would try a different school for his/her son/daughter. The availability of textbooks at home seems to matter for lowering the risk of leaving school for females.

Turning the attention to the role of the labour market on the conditional probability of dropping out, the results indicate that female youth unemployment rates do not play a significant role in explaining differences on school dropout hazard rates. This is consistent with the preliminary findings shown in Section 2. That labour demand seems to affect the schooling decisions of males but not those of women is found in the empirical work for developed countries. Rice (2001), Petrongolo and San Segundo (2002), Clark (2011).

As in the case of males, I reject the hypothesis that each person's hazard rate does not change over time. The shape of female baseline hazard seems to increase over time during the first three years, decelerate at the fourth and fifth years and increases again in the last two years.

3.4.3 Extensions

Models presented above were re-estimated allowing the existence of unobserved heterogeneity in the hazard. Specifically, the discrete time Proportional Hazard model with Gamma heterogeneity was estimated[38].

Table 3.6 presents the results for males using a similar specification that of Table 3.3, column (2). Estimates for other specifications are not shown because results were similar.

The LR test indicates that the model with unobservable heterogeneity is not statistically significant. The hypothesis that Gamma var is zero cannot be rejected, therefore frailty is unimportant. As discussed by Jenkins (2005), it might be less biased of excluding unobserved heterogeneity when, as in my case, a fully flexible specification for the baseline hazard function

[38]The estimations were performed using pgmhaz8 command in Stata developed by Jenkins (1997). The command m1 deriv0 is used, beginning with an initial value beta estimated in the discrete time PH cloglog model. See STB-39 for more details.

TABLE 3.5

Estimates for the Discrete Time Proportional Hazard Model. Females.

	(1)	(2)	(3)	(4)
Incomp. primary or less (parents)	1.966***	1.992***	1.950***	1.969***
	(0.478)	(0.480)	(0.477)	(0.479)
Completed primary or incomp. secondary (parents)	0.829**	0.861**	0.808**	0.837**
	(0.401)	(0.401)	(0.402)	(0.402)
Age first time enrolled in secondary educ.	0.507***	0.497***	0.562***	0.555***
	(0.163)	(0.163)	(0.176)	(0.176)
Repeat at least once in primary	0.327	0.366	0.351	0.394
	(0.491)	(0.495)	(0.492)	(0.495)
School located in Conurbano	0.165	0.189	0.174	0.202
	(0.315)	(0.316)	(0.315)	(0.315)
Half-time education	0.421	0.453	0.436	0.474
	(0.471)	(0.474)	(0.474)	(0.479)
Never changed sec. school	0.974***	0.971**	0.988***	0.989***
	(0.375)	(0.377)	(0.375)	(0.377)
Bilingual education	-0.322	-0.304	-0.333	-0.318
	(0.405)	(0.404)	(0.405)	(0.403)
Private (religious) educ.	-0.734	-0.776	-0.720	-0.761
	(0.493)	(0.493)	(0.494)	(0.494)
Private (non-religious) educ,	-0.669	-0.682	-0.610	-0.620
	(0.495)	(0.494)	(0.499)	(0.497)
Vocational education	0.852	0.904	0.881	0.921
	(0.641)	(0.647)	(0.644)	(0.650)
Books available at home	-0.754**	-0.776**	-0.786**	-0.817**
	(0.331)	(0.332)	(0.327)	(0.329)
Youth (16-18), unskilled unemployment rate	-1.017	-1.189		
	(1.243)	(1.251)		
Youth (18-21), unskilled unemployment rate			-2.276	-2.487
			(1.811)	(1.821)
Failure in Year t		-0.680		-0.692
		(0.616)		(0.615)
Baby boom	-0.550	-0.596*	-0.824*	-0.884*
	(0.353)	(0.356)	(0.455)	(0.459)
Baseline Hazard (BH)				
Year 1	-11.18***	-11.01***	-11.50***	-11.36***
	(2.379)	(2.385)	(2.428)	(2.436)
Year 2	-10.92***	-10.64***	-11.22***	-10.99***
	(2.365)	(2.376)	(2.412)	(2.423)
Year 3	-10.12***	-9.849***	-10.42***	-10.19***
	(2.345)	(2.358)	(2.391)	(2.405)
Year 4	-10.41***	-10.17***	-10.70***	-10.52***
	(2.345)	(2.354)	(2.392)	(2.402)
Year 5	-10.37***	-10.18***	-10.67***	-10.52***
	(2.337)	(2.343)	(2.385)	(2.395)
Year 6	-8.980***	-8.818***	-9.301***	-9.181***
	(2.318)	(2.325)	(2.371)	(2.381)
Year 7	-9.740***	-9.583***	-10.06***	-9.948***
	(2.520)	(2.528)	(2.558)	(2.569)
Ho: BH is constant (Chi2(7))	38.66	37.16	39.08	37.76
Prob > chi2 =	0.00	0.00	0.00	0.00
Observations	1,483	1,483	1,483	1,483
DF	21	22	21	22
Chi2	449.0	446.7	446.7	444.2

Note: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Source: Author's calcluation using EJJ data.

is fitted.

TABLE 3.6
Estimates for the Discrete Model with Unobserved Heterogeneity. Males

Variables	Coef.	Srd.Err.	z
Completed primary or incomp.secondary (parents)	0.438	0.556	0.79
Age first time enrolled in secondary educ.	1.003	0.410	2.44
Age first time enrolled in secondary educ.	0.456	0.199	2.29
Repeat at least once in primary	0.474	0.502	0.94
School located in Conurbano	1.197	0.497	2.41
Half-time education	-0.004	0.455	-0.01
Never changed sec. school	0.172	0.385	0.45
Bilingual education	-0.770	0.705	-1.09
Private (religious) educ.	-1.435	0.691	-2.08
Private (non-religious) educ,	-0.406	0.677	-0.6
Vocational education	-0.283	0.493	-0.57
Books available at home	-0.409	0.414	-0.99
Baby boom	-1.181	0.456	-2.59
Youth (16-18), unskilled unemployment rate	-4.471	1.960	-2.28
Failure_t	-0.480	0.525	-0.92
Year 1	-9.063	2.806	-3.23
Year 2	-8.331	2.703	-3.08
Year 3	-7.771	2.593	-3
Year 4	-7.571	2.522	-3
Year 5	-7.229	2.445	-2.96
Year 6	-6.752	2.444	-2.76
Year 7	-6.102	2.403	-2.54
ln_varg			
_cons	0.059	1.072	0.05
Gamma var	1.061	1.136	0.93
LR test of Gamma var==0: chibar2(01)=1.28911	Prob.>=chibar2=		0.13

Source: Author's calcluation using EJJ data.

The model with unobserved heterogeneity for females were not estimated due to problems in convergence. According to Jenkins (2005), this could have happened because the Gamma variance is very close to zero.

Finally, as suggested in the literature, see Allison (1982), where original models from Table 3.3 were re-estimated using unordered multinomial choice specification (not shown)- multinomial logit for the competing risk by distinguishing among different kinds of events: dropout, graduation and right censured. Main conclusions remained unchanged.

3.5 Final Remarks

In this paper, the role of labour market conditions on youth schooling decisions was examined using discrete time proportional hazard models.

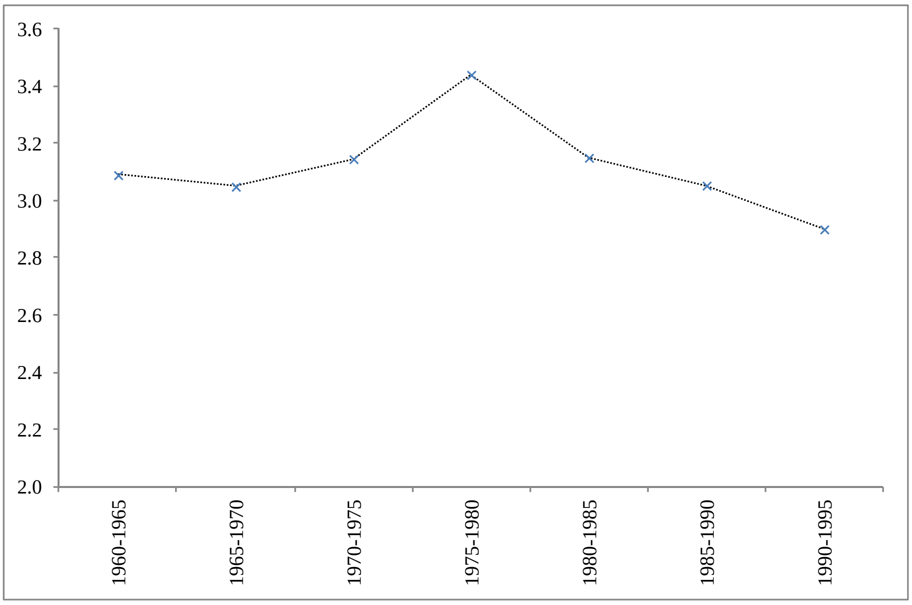
The results indicated that - for males - youth unemployment rate has a significant effect on the hazard rate for dropping out of school. Young people who faced low unemployment rates by the legal working age were more likely to leave school than those who experienced worse conditions in the labour market. These results are consistent with the human capital theory.

For the case of females, the labour market does not have a significant impact on the hazard rate of dropping out. However, parental education and age at entry showed to strongly affect the risk of leaving schooling not only for females but also for males.

The analysis of the shape of the baseline hazard rate suggested that the conditional probability of dropping out is not constant across the school years, but instead increases for males, and to some extent for females.

3.6 Appendix

FIGURE 3.3
Global Fertility Rate. Total Argentina. 1960-1995



Source: INDEC.

References

1. Allison, P. D. (1982). Discrete-time methods for the analysis of event histories en S. Leinhardt (ed.) Sociological Methodology 1982. Jossey-Bass Publishers, San Francisco, pp. 61-97.

2. Binstock, G. and Cerrutti, M. (2005). Carreras Truncadas. El abandono escolar en el nivel medio en la Argentina. UNICEF/ Fondo de las Naciones Unidas para la Infancia. Buenos Aires, Argentina.

3. Cox, D.R. (1972) Regression Models and Life tables, (with discussion) Journal of the Royal Statistical Society, 34: 187-220

4. Cox D. R y Oakes D. (1984). Analysis of Survival data. London: Chapman and Hall.

5. Dex, S. (1991) The Reliability of Recall Data : A Literature Review (ESRC working paper, No. 11). Colchester: University of Essex.

6. Eisenhower, D., Mathiowetz, N. A. and Morganstein, D. (2004) Recall Error: Sources and Bias Reduction Techniques, in Measurement Errors in Surveys (eds P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz and S. Sudman), John Wiley & Sons, Inc., Hoboken, NJ, USA. doi: 10.1002/9781118150382.ch8

7. INDEC (2010). Estimaciones y proyecciones de población. Total del país. 1950-2015. Población por sexo y grupos quinquenales de edad. Total del país. Años 1950-2015

8. INDEC (2005). Informe sobre la no respuesta de los hogares y la población objetivo en la aplicación del modulo "Educación y Empleo de los jóvenes". DIRECCIÓN ENCUESTA PERMANENTE DE HOGARES- Departamento de Muestreo.

9. Jenkins, S. P, (1997). sbe 17 Discrete time proportional hazards regression (pgmhaz) Stata Technical Bulletin STB-39, September 1997, 22-32. Reprinted in Stata Technical Bulletin Reprints, vol. 7, H.J. Newton (ed.), 109–121, 1998. College Station TX: Stata Corporation. http://www.statapress.com/journals/stbcontents/stb39.pdf

10. Jenkins, S.P. (1995). Easy estimation methods for discrete-time duration models. Oxford Bulletin of Economics and Statistics 57 (1): 129-138.

11. Kiefer, N (1988). Economic duration data and hazard functions. Journal of Economic Literature, Vol 26, pp. 646-679.

12. Klein, J. y Moeschberger, M. (1997). Survival Analysis of Failure Data. New York: Wiley.

13. Lancaster, T (1990). The econometric analysis of transition data. (1ra ed.). Cambrige University Press.

14. Marchionni, M. y Sosa Escudero, W. [2000], sin fecha, "Los determinantes de la decisión de escolarización", Universidad Nacional de La Plata.

15. Mathiowetz, Nancy A. , Charlie Brown, and John Bound (2001) "Measurement Error Issues in Surveys of the Low Income Population." Data Collection on Low Income and Welfare Populations. Washington, D.C.: National Academy Press

16. Meyer, B. D. (1990). Unemployment insurance and unemployment spells. Econométrica 58(4): pp. 757-782.

17. Micklewright, J., Pearson, M., and Smith S., (1990), "Unemployment and Early School Leaving", The Economic Journal, 100(400), 163-169. pages 163-69, Supplemen.

18. Paz, Jorge, 2008, "Cambios en el ingreso del hogar y sus efectos sobre la escolaridad de menores (Argentina 1995-2003)", Salta, Universidad Nacional de Salta, documento de trabajo del IELDE, año 2008, núm. 2, p. 1-44.

19. Petrongolo, B. and San Segundo, M. J. (2002), "Staying-on at school at 16: the impact of labor market conditions in Spain", Economics of Education Review, 21, 353–365.

20. Prentice R. L. & Gloeckler (1978). Regression analysis of grouped survival data with application to breast cancer data. Biometrics 34, pp.57-67

21. Rice, P. (1999), "The Impact of Local Labour Markets on Investment in Further Education: Evidence from the England and Wales Youth Cohort Studies", Journal of Population Economics, 12(2), 287-312

22. Singer and Willet (1991). "From whether to when: New methods for studying student dropout and teacher attrition". Review of Educational Research, 61 (4), pp. 407-450.